

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: May 12, 2003, 10:09:12 ; Search time 78 Seconds
(without alignments)
736.295 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 2275

Sequence: 1 MHHHHHSSGLVPRGSMKE.....TCECTKPSYPLFDGIFCSS 431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :			
A_Geneseq_101002:*			
1:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*		
2:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*		
3:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*		
4:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*		
5:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*		
6:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*		
7:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*		
8:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*		
9:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*		
10:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*		
11:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*		
12:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*		
13:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*		
14:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*		
15:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*		
16:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*		
17:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*		
18:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*		
19:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*		
20:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*		
21:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*		
22:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*		
23:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1970	86.6	375	22	AA83926
2	1175	51.6	1639	19	AAW54145
3	1161.5	51.1	355	20	AAV09372
4	1161.5	51.1	355	20	AAV05832
5	1161.5	51.1	361	20	AAV09373
6	1161.5	51.1	361	20	AAV05833
7	1154.5	50.7	376	20	AAV09374
8	1154.5	50.7	376	20	AAV05834
9	1007.5	44.3	1654	6	AAV50777
10	640.5	28.2	262	22	AA837610

11	527	23.2	116	18	AAW36103
12	527	23.2	116	18	AAW22592
13	527	23.2	127	18	AAW22593
14	527	23.2	127	18	AAW36102
15	525	23.1	108	22	AA837609
16	524	23.0	96	22	AA837608
17	406	17.8	206	22	AAU00669
18	302	13.3	402	23	ABG67266
19	301	13.2	407	23	ABG67270
20	297	13.1	407	23	ABG67267
21	294	12.9	54	14	AA841357
22	283.5	12.5	1807	22	AA885697
23	283.5	12.5	2028	22	AA885698
24	282	12.4	49	14	AA841355
25	280	12.3	108	22	AA847487
26	279	12.3	49	14	AA841354
27	278.5	12.2	289	22	AAE15033
28	278.5	12.2	289	23	AA847732
29	278.5	12.2	824	23	AA848247
30	278	12.2	54	14	AA841356
31	277.5	12.2	692	21	AA811532
32	275	12.1	467	21	AA85150
33	275	12.1	467	21	AA85150
34	274	12.0	76	21	AA826110
35	274	12.0	167	22	AA85253
36	274	12.0	167	22	AA85254
37	274	12.0	551	22	AA85251
38	274	12.0	551	22	AA85252
39	272	12.0	93	22	AA837611
40	263.5	11.6	138	21	AA811534
41	259.5	11.4	408	20	AAW84315
42	256.5	11.3	106	14	AA841358
43	255.5	11.2	284	23	AAE20460
44	250	11.0	59	23	AB80661
45	250	11.0	218	21	AA811533

ALIGNMENTS

RESULT 1	
AA83926	AA83926 standard; Protein; 375 AA.
ID	AA83926 standard; Protein; 375 AA.
XX	
AC	AA83926;
XX	
DT	23-JUL-2001 (first entry)
XX	
DE	A major merozoite surface protein-1 fragment of 42kDa.
XX	
KW	Major merozoite surface protein-1; MSP1-42; melittin signal peptide;
KW	malaria vaccine
XX	
OS	Plasmodium falciparum.
XX	
PN	WO200134188-A1.
XX	
PD	17-MAY-2001.
XX	
PP	09-NOV-2000; 2000WO-US31064.
XX	
PR	12-NOV-1999; 99US-0165178.
PR	01-DEC-1999; 99US-0168327.
PR	22-AUG-2000; 2000US-0226861.
XX	
PA	(UYHA-) UNIV HAWAII.
PA	(UYCH-) UNIV CHINESE HONG KONG.
PA	(QUEE-) QUEEN EMMA FOUND.
XX	
PI	Hui GSN, Lap-Yin P, Ho WKK;
XX	
DR	WPI; 2001-335879/35.
DR	N-PSDB; AAF89840.

XX Producing malaria vaccine, useful for treatment or prevention of all
PT forms of malaria in humans, by expressing immunogenic merozoite protein
PT fragment in a baculovirus system -
XX
XX Example 3; Page 87-88; 95pp; English.
XX
XX The present sequence represents a major merozoite surface protein-1
CC C-terminal fragment of 42kDa (MSP-1-42). This fragment is linked to a
CC melittin signal peptide, and then expressed in a in a
CC silkworm/baculovirus system. The protein is used to prepare a
CC malaria vaccine, which is used to treat or prevent malaria, caused by
CC any of the four species of Plasmodium that infect humans.
XX
XX Sequence 375 AA;
XX
XX Query Match 86.6%; Score 1970; DB 22; Length 375;
XX Best Local Similarity 100.0%; Pred. No. 2e-132;
XX Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 57 MAISVTMDNLSGFGFENEVDVYLYPLAGVYRSLKQIEKNFTFNLDNLSRLKRRK 116
XX Db 1 MAISVTMDNLSGFGFENEVDVYLYPLAGVYRSLKQIEKNFTFNLDNLSRLKRRK 60
XX
XX QY 117 YFLOVLESDLMQFHHISSNEVYIIEDSFKLLNSEQNTLLSKYIKESVENDIKFAQEG 176
XX Db 61 YFLOVLESDLMQFHHISSNEVYIIEDSFKLLNSEQNTLLSKYIKESVENDIKFAQEG 120
XX
XX QY 177 SYEYKVLAKYKDDLESIKKVKIEKEKPPSPPTPPSPAKTDEQKESKFLPLTNIET 236
XX Db 121 SYEYKVLAKYKDDLESIKKVKIEKEKPPSPPTPPSPAKTDEQKESKFLPLTNIET 180
XX
XX QY 237 LYNNLVNKIDYLYNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIK 296
XX Db 181 LYNNLVNKIDYLYNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIK 240
XX
XX QY 297 KLINDTKKMLGKLLSTGLVQNPNTTISKLEGFODMLNISOHCQVKKQCPENSGCF 356
XX Db 241 KLINDTKKMLGKLLSTGLVQNPNTTISKLEGFODMLNISOHCQVKKQCPENSGCF 300
XX
XX QY 357 RHLDERECKCLLNYKQGDCKVCNPNPTCNENNGCGDADATCTEEDSGSRKKITCECT 416
XX Db 301 RHLDERECKCLLNYKQGDCKVCNPNPTCNENNGCGDADATCTEEDSGSRKKITCECT 360
XX
XX QY 417 KPDSPYPLFDGIFCSS 431
XX Db 361 KPDSPYPLFDGIFCSS 375
XX
XX RESULT 2
XX AAW54145
XX ID AAW54145 standard; Protein; 1639 AA.
XX
XX AC AAW54145;
XX
XX XX
XX 23-SEP-1998 (first entry)
XX
XX DE P. falciparum synthetic gp190 protein.
XX
XX gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
XX monoclonal antibody; passive immunisation; parasite.
XX
XX Plasmodium falciparum.
XX Synthetic.
XX
XX WO9814583-A2.
XX
XX 09-APR-1998.
XX
XX 02-OCT-1997; 97WO-EP05441.
XX
XX 02-OCT-1996; 96DE-4040817.
XX

(BUJA/) BUJARD H.
XX Bujard H, Pan W, Tolle R;
XX WPI: 1998-240088/21.
XX N-PSDB; AAV21451, AAV35363.
XX
XX Recombinant production of complete gp190/MSP-1 Plasmodium surface
XX protein - useful in anti-malaria vaccines, also stabilising genes by
XX reducing their AT content
XX
XX Example 1; Fig 3c; 48pp; German.
XX
XX This sequence represents a modified Plasmodium falciparum gp190/MSP-1
XX (merozoite surface) protein. The gene encoding this protein has been
XX stabilised by reducing the AT content of the nucleotide sequence. Such a
XX protein is useful in vaccines against malaria or for producing monoclonal
XX antibodies (for passive immunisation). The complete gp190 protein can now
XX be produced outside the parasite and has, at least over extended regions,
XX the native pattern of folding. Larger amounts of the protein can be
XX produced recombinantly than would be possible using the parasites as
XX source.
XX
XX Sequence 1639 AA;
XX
XX Query Match 51.6%; Score 1175; DB 19; Length 1639;
XX Best Local Similarity 56.6%; Pred. No. 7.3e-75;
XX Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;
XX
XX QY 34 PDLGTDGDDKAMADIGSIEGRGTWAIISVTMDNLSGFGFENEVDVYLYPLAGVYRSLKQI 93
XX Db 1244 PIFGESEED--YDGLGQVVTGEAVTPSV-IDNLSKIENEYEVLYLPLAGVYRSLKQI 1300
XX
XX QY 94 EKNFTFNLDNLSRLKRRKYFLDVLESDLMQFHHISSNEVYIIEDSFKLLNSEQNT 153
XX Db 1301 ENNVYTFNVNVKIDLSNFRFNKRENFKNVLESDDLIPYKDLTSSNVVVKDPYFLKKEKRD 1360
XX
XX QY 154 LLKSYKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVKIEKEKPPSPPTTPP 213
XX Db 1361 FLSSYNTIKDSIDIDINFANDVLGYIKLSEKYSKSLDSIKKYI----- 1404
XX
XX QY 214 SPAKTDEQKESKFLPLTNIETLYNNLVNKIDYLYNLKAKINDCNVEKDEAHVKITKL 273
XX Db 1405 ----NDKQGENEKYLPFLNNTIETLYNNLVNKIDYLYNLKAKINDCNVEKDEAHVKITKL 273
XX
XX QY 274 SLDKAIIDDKIDLFKNPYDFAIKKLLINDTKKMLGKLLSTGLV-QNFPNTIISKLEGG 332
XX Db 1461 NYLKTIOQKLADFPKNNFVGIADLSTDYNNHNLTLTFLSTGMVFENLAKTVLSNLLDGN 1520
XX
XX QY 333 FQDMLNISOHCQVKKQCPENSGCFRHLDERECKCLLNYKQGDCKVCNPNPTCNENNGG 392
XX Db 1521 LQGLMLNISOHCQVKKQCPENSGCFRHLDERECKCLLNYKQGDCKVCNPNPTCNENNGG 1580
XX
XX QY 393 CDADATCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431
XX Db 1581 CDADATCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 1619
XX
XX RESULT 3
XX AAY09372
XX ID AAY09372 standard; Protein; 355 AA.
XX
XX AC AAY09372;
XX
XX XX
XX 31-AUG-1999 (first entry)
XX
XX DE Merozoite surface protein MSP-1-42.
XX
XX MSP-1; merozoite surface protein; malaria; vaccine;
XX protein engineering; protein expression; codon usage;
XX transgenic animal.
XX
XX Plasmodium falciparum.
XX

Db 127 SEKYKSDLSKYYI-----NDKQGENEKYLPFLNNIETLYKTVN 166
QY 243 NKIDYILINKAKTNCNVEKDEAHVITKILSLDKATDDKIDLFKNPYDFEAIKKLND 302
Db 167 DKIDLFVHLEAKVNTYKSNVEVAKELNLYKTQDKLADFKKNNFVGIADLSTDY 226
QY 303 TKKDMGLKLLSTGLV-QNFPNTIISKIEGKFQDMLNISQHCYKQCPENSGCFRHLDE 361
Db 227 NHNLLTKFLSTGMVFENLAKTVLSNLLDGNLQGLMNLISQHCYKQCPENSGCFRHLDE 286
QY 362 REECKCLLNYKQEGDKCVENPNTCNENNGCGDADATCTEEDSGSSRKKITCECTKPDYSY 421
Db 287 REECKCLLNYKQEGDKCVENPNTCNENNGCGDADAKCTEEDSGSNGKKITCECTKPDYSY 346
QY 422 PLFDGIFCS 430
Db 347 PLFDGIFCS 355
RESULT 5
AAAY09373
ID AAY09373 standard; Protein; 361 AA.
XX
AC AAY09373;
XX
DT 31-AUG-1999 (first entry)
XX
DE Merozoite surface protein MSP-1-42.
XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX
OS Plasmodium falciparum.
XX
PN WO9920774-A2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22226.
XX
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GENZ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI; 1999-288313/24.
DR P-PSDB; AAX56009.
XX
PT Modified malarial protein for use in anti-malarial vaccines
XX
PS Example; Fig 2; 35pp; English.
XX
CC This present sequence comprises a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The C-terminal end of the sequence is
CC modified to include a 6xHis tag. A nucleic acid (see AAX56008)
CC encoding MSP-1-42 has been modified according to a method
CC of the invention in order to improve expression in mammalian cells
CC and in transgenic animals by reducing the AT content and removing
CC mRNA instability motifs. The invention allows expression of
CC MSP-1-42 in the milk of transgenic animals, and also provides a DNA
CC vaccine comprising a vector containing the altered MSP-1-42 nucleic
CC acid.
XX
SQ Sequence 361 AA;
Query Match 51.1%; Score 1161.5; DB 20; Length 361;
Best Local Similarity 59.1%; Pred. No. 9e-75;

Matches 218; Conservative 54; Mismatches 76; Indels 21; Gaps 2;
QY 63 MDNLTSGFENEYDVIYKPLAGVYRSRKQIEKNIETFNLDILNSRLKRRKYFELDVL 122
Db 7 IDNLSIKIENEYEVLYKPLAGVYRSRKQLENNVMTFNVKDILNSRFNKRFRNVL 66
QY 123 ESOLMOPKHSSNYIIDSFKLNSQKNTLLSKYIKRESVENDIKFAOEGISYIEKV 182
Db 67 ESDLIPYKDLTSSNVYKDPVKFLNKEKRDKFLSSYINIKSDIDTDFANDVLGYKIL 126
QY 183 LAKYKDDLESITKKVKEEKFPSPPTTPPSAKTDEOKKESKFLPLNIETLYNNLV 242
Db 127 SEKYKSDLSITKKYI-----NDKQGENEKYLPFLNNIETLYKTVN 166
QY 243 NKIDYILINKAKTNCNVEKDEAHVITKILSLDKATDDKIDLFKNPYDFEAIKKLND 302
Db 167 DKIDLFVHLEAKVNTYKSNVEVAKELNLYKTQDKLADFKKNNFVGIADLSTDY 226
QY 303 TKKDMGLKLLSTGLV-QNFPNTIISKIEGKFQDMLNISQHCYKQCPENSGCFRHLDE 361
Db 227 NHNLLTKFLSTGMVFENLAKTVLSNLLDGNLQGLMNLISQHCYKQCPENSGCFRHLDE 286
QY 362 REECKCLLNYKQEGDKCVENPNTCNENNGCGDADATCTEEDSGSSRKKITCECTKPDYSY 421
Db 287 REECKCLLNYKQEGDKCVENPNTCNENNGCGDADAKCTEEDSGSNGKKITCECTKPDYSY 346
QY 422 PLFDGIFCS 430
Db 347 PLFDGIFCS 355
RESULT 6
AAAY05833
ID AAY05833 standard; Protein; 361 AA.
XX
AC AAY05833;
XX
DT 02-AUG-1999 (first entry)
XX
DE Merozoite surface protein MSP-1-42.
XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX
OS Plasmodium falciparum.
XX
PN WO9920766-A2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22225.
XX
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GENZ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI; 1999-302742/25.
DR N-PSDB; AAX25587.
XX
PT New modified recombinant nucleic acid sequences useful for producing
PT malarial DNA vaccine
XX
PS Disclosure; Fig 2; 43pp; English.
XX
CC This present sequence comprises a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The C-terminal end of the sequence is
CC modified to include a 6xHis tag. Nucleic acids (see AAX25586 and

QY 37 GTDDDDKAMADIGSIEGRGTMATISVTMDNLISGFENEDYVILYKPLAGVYRSLLKQIKEN 96
DB 1261 GESEED--YDDLGOVVTGEAVTTSV-IDNLISKIENEYEVLYLKPLAGVYRSLLKQLENN 1317
QY 97 IITFNLNLDILNSRLKRRKYFLVDLESDLQMFKHSSNEYIIEDSFKLLNSEQKNTLLK 156
DB 1318 VMTFNVNVDILNSRFNKNENKFNVLSDLPYKOLTSNVYVKDPYKFLNKEKRDKFLS 1377
QY 157 SYKIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEEKEKFPSSPPPTPPSPA 216
DB 1378 SYNYIKDSIDTDFINFANDVGLYKILSEYKSDLSIKKYI----- 1418
QY 217 KTDEQKESKFLPFLNIETLYNNVNLKIDDLINLAKINDCNVEKDEAHVKITKLSDL 276
DB 1419 -NDKOGENEKYLPLFNLIETLYKTVNDKIDLFVHLEAKVNLTYEKSNNVEVKIKELNVL 1477
QY 277 KATDDKIDLFKNPYDFAEAKKLINDTKDMLGKLLSTGLV-QNFPNTIISKIEGKF-Q 334
DB 1478 KTIQDKLADFKNNNNEVGIADLSTDYNNHNNLTFLSTGVMFENLKSLSNLDWKLAR 1537
QY 335 DMLNISQHCVRKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCD 394
DB 1538 YVKHFTTPHRKMTMIQSSGCFRHLDERECKCLLNYKQEGSKCVENSNPTCNENNGGCD 1597
QY 395 ADATCEDSGSGSRKKTCECTKPDSPYPLFDGIFCSS 431
DB 1598 ADAKTEDSGSGSKKTCECTKPCYPLSMVIFCSS 1634
RESULT 10
AAB37610
ID AAB37610 standard; Protein; 262 AA.
AC AAB37610;
XX
DT 27-FEB-2001 (first entry)
XX
DE Merozoite surface protein-133.
XX
KW Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS Plasmodium falciparum.
XX
PN WO200063245-A2.
XX
PD 26-OCT-2000.
XX
PF 20-APR-2000; 2000WO-CB01558.
XX
PR 20-APR-1999; 99GB-0009072.
PR 13-MAY-1999; 99US-0311817.
PR 25-MAY-1999; 99CA-2271451.
XX
XX (MEDI-) MEDICAL RES COUNCIL.
PA
PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaipibull C;
XX
XX WPI; 2001-015762/02.
DR N-PSDB; AAC68978.
XX
XX Novel variants of the C-terminal fragment of Plasmodium merozoite
PT surface protein-1, useful as vaccines for treating or preventing
PT malaria.
XX
XX Example 5; Fig 15; 126pp; English.
PS
XX
XX The present invention relates to non-natural variants of a C-terminal
CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1.4.2, compared to natural MSP-1.1.9. The non-natural variants of the

CC present invention are useful for immunising a mammal against malaria, and
CC can be used to treat malaria. The present sequence is MSP-133 protein.
XX
SQ Sequence 262 AA;
Query Match 28.2%; Score 640.5; DB 22; Length 262;
Best Local Similarity 46.7%; Pred. No. 8.7e-38;
Matches 129; Conservative 51; Mismatches 75; Indels 21; Gaps 2;
QY 63 MDNLISGFENEDYVILYKPLAGVYRSLLKQIKENIETFLNLDILNSRLKRRKYFLDVL 122
DB 7 IDNLISKIENEYEVLYLKPLAGVYRSLLKQLENNVMTFNVNVDILNSRFNKNENKFNVL 66
QY 123 ESLLMDFKHSSNEYIIEDSFKLLNSEQKNTLLKSYKIKESVENDIKFAQEGISYVEKV 182
DB 67 ESDLIPYKOLTSNVYVKDPYKFLNKEKRDKFLSSYNIKDSIDTDFINFANDVGLYKIL 126
QY 183 LAKYKDDLESIKKVIKEEKEKFPSSPPPTPPSPAKTDEQKESKFLPFLNIETLYNNLV 242
DB 127 SEKYKSDLSIKKYI-----NDKOGENEKYLPLFNLIETLYKTVN 166
QY 243 NKTDYDLINLAKINDCNVEKDEAHVKITKLSDLKATDDKIDLFKNPYDFAEAKKLIND 302
DB 167 DKIDLFVHLEAKVQLTYEKSNNVEVKIKELNYLKTIDQDKLADFKNNNNEVGIADUSTDY 226
QY 303 TKKDMLGKLLSTGLV-QNFPNTIISKIEGKFQDML 337
DB 227 NHNNLTFLSTGVMFENLAKTVLSNLDGNGML 262
RESULT 11
AAB36103
ID AAB36103 standard; Protein; 116 AA.
XX
AC AAB36103;
XX
DT 25-MAR-1998 (first entry)
XX
DE PfMSP1(p19)A protein sequence.
XX
KW Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS Plasmodium falciparum.
OS Synthetic.
FH Key Location/Qualifiers
FT Region 1..95 /note= "amino acids derived from P. falciparum MSP1 p19
FT fragment"
FT Region 96..116 /note= "glycosylphosphatidylinositol anchoring sequence"
XX
PN WO9730158-A2.
XX
PD 21-AUG-1997.
XX
PF 14-FEB-1997; 97WO-FR00290.
XX
PR 14-FEB-1996; 96FR-0001822.
XX
XX (INSP) INST PASTEUR.
PA (UUNY) UNIV NEW YORK STATE.
XX
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
PI Roth C;
XX
XX WPI; 1997-425033/39.
DR N-PSDB; AAT94550.
XX
XX Recombinant protein containing the merozoite surface protein-1 p19
PT fragment - useful in anti-malarial vaccines, diagnosis and protein
PT purification

XX PS Disclosure; Fig 1B; 85pp; French.

XX CC This is the amino acid sequence of a recombinant protein comprising

CC amino acids 1613-1705 of the Plasmodium falciparum merozoite surface

CC protein 1 (MSPI) 19 kD C-terminal fragment (p19), linked to a

CC glycosylphosphatidylinositol membrane anchoring sequence. p19 is the

CC C-terminal fragment of the 42 kD MSPI from Plasmodium species.

CC The recombinant protein can be used for the production of anti-malarial

CC vaccines, where the p19 fragment provides a high level of protective

CC immunity since it includes epitopes not presented in the p42 fragment.

XX Sequence 116 AA;

XX Query Match 23.2%; Score 527; DB 18; Length 116;

XX Best Local Similarity 96.8%; Pred. No. 3.7e-30;

XX Matches 91; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 338 NISQHCYKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENGGCDADA 397

DB 3 NISQHCYKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENGGCDADA 62

QY 398 TCTEEDSGSRKKITCECTKPDSPFLFDGIFCSS 431

DB 63 KCTEEDSGSGKKITCECTKPDSPFLFDGIFCSS 96

RESULT 12

AAW22592

ID AAW22592 standard; Protein; 116 AA.

XX AC AAW22592;

XX DT 25-MAR-1998 (first entry)

XX DE PfMSPI(p19)A protein sequence.

XX KW Plasmodium vivax; merozoite surface protein; MSPI; p19;

XX KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.

XX OS Plasmodium falciparum.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Region 1..95

FT /note= "amino acids derived from P. falciparum MSPI p19 fragment"

FT Region 96..116

FT /note= "glycosylphosphatidylinositol anchoring sequence"

XX WO9730159-A2.

XX 21-AUG-1997.

XX 14-FEB-1997; 97WO-FR00291.

XX 14-FEB-1996; 96FR-0001821.

XX (INSP) INST PASTEUR.

XX (UYNV) UNIV NEW YORK STATE.

XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;

XX Roth C;

XX WPI; 1997-425034/39.

XX P-PSDB; AAW22592.

XX Recombinant protein containing Plasmodium merozoite surface

XX protein-l p42 fragment - useful in antimalarial vaccines, also new

XX antibodies for diagnosis and protein purification

XX PS Disclosure; Fig 1B; 85pp; French.

CC This is the amino acid sequence of a recombinant protein comprising

CC amino acids 1613-1705 of the Plasmodium falciparum merozoite surface

CC protein 1 (MSPI) 19 kD C-terminal fragment (p19), linked to a

CC glycosylphosphatidylinositol membrane anchoring sequence. p19 is the

CC C-terminal fragment of the 42 kD MSPI from Plasmodium species.

CC The recombinant protein can be used for the production of anti-malarial

CC vaccines, where the p19 fragment provides a high level of protective

CC immunity since it includes epitopes not presented in the p42 fragment.

XX Sequence 116 AA;

XX Query Match 23.2%; Score 527; DB 18; Length 116;

XX Best Local Similarity 96.8%; Pred. No. 3.7e-30;

XX Matches 91; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 338 NISQHCYKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENGGCDADA 397

DB 3 NISQHCYKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENGGCDADA 62

QY 398 TCTEEDSGSRKKITCECTKPDSPFLFDGIFCSS 431

DB 63 KCTEEDSGSGKKITCECTKPDSPFLFDGIFCSS 96

RESULT 13

AAW22593

ID AAW22593 standard; Protein; 127 AA.

XX AC AAW22593;

XX DT 25-MAR-1998 (first entry)

XX DE PfMSPI(p19)S protein sequence.

XX KW Chimeric; Plasmodium vivax; merozoite surface protein; MSPI; p19;

XX KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.

XX OS Chimeric - Plasmodium vivax.

XX OS Chimeric - Plasmodium falciparum.

XX FH Key Location/Qualifiers

FT Peptide 1..19

FT /note= "signal peptide"

FT Protein 20..127

FT /note= "mature protein"

FT Region 1..32

FT /note= "derived from P. vivax MSPI"

FT Region 33..34

FT /note= "encoded by restriction enzyme sequence used to create the chimeric sequence"

FT Region 35..127

FT /note= "derived from P. falciparum C-terminal p19 fragment of MSPI"

XX WO9730159-A2.

XX 21-AUG-1997.

XX 14-FEB-1997; 97WO-FR00291.

XX 14-FEB-1996; 96FR-0001821.

XX (INSP) INST PASTEUR.

XX (UYNV) UNIV NEW YORK STATE.

XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;

XX Roth C;

XX WPI; 1997-425034/39.

XX P-PSDB; AAW22592.

XX Recombinant protein containing Plasmodium merozoite surface

XX protein-l p42 fragment - useful in antimalarial vaccines, also new

PT antibodies for diagnosis and protein purification

XX Disclosure; Fig 1C; 85pp; French.

CC This is the amino acid sequence of a chimeric protein comprising amino acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1) linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.

CC The recombinant protein can be used for the production of anti-malarial vaccines, where the p19 fragment provides a high level of protective immunity since it includes epitopes not presented in the p42 fragment.

XX Sequence 127 AA;

Query Match 23.2%; Score 527; DB 18; Length 127;

Best Local Similarity 94.8%; Pred. No. 4.1e-30;

Matches 91; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 335 DMLNISQHCVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENGGCD 394

DB 32 DEFNISQHCVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENGGCD 91

QY 395 ADATCTEEDSGSRKKITCECTKPDSPYPLFDGIFCS 430

DB 92 ADAKTEEDSGSGNGKKTCECTKPDSPYPLFDGIFCS 127

RESULT 14

AAW36102

ID AAW36102 standard; Protein; 127 AA.

XX AAW36102;

XX 25-MAR-1998 (first entry)

XX PfMSP1(p19)S protein sequence.

XX Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;

XX Plasmodium falciparum; malaria; vaccine; immunity; epitope.

XX Chimeric - Plasmodium vivax.

XX Chimeric - Plasmodium falciparum.

XX Key Location/Qualifiers

FT Peptide 1..19

FT /note= "signal peptide"

FT Protein 20..127

FT /note= "mature protein"

FT Region 1..32

FT /note= "derived from P. vivax MSP1"

FT Region 33..34

FT /note= "encoded by restriction enzyme sequence used to create the chimeric sequence"

FT Region 35..127

FT /note= "derived from P. falciparum C-terminal p19 fragment of MSP1"

XX WO9730158-A2.

XX 21-AUG-1997.

XX 14-FEB-1997; 97WO-FR00290.

XX 14-FEB-1996; 96FR-0001822.

XX (INSP) INST PASTEUR.

XX (UUNY) UNIV NEW YORK STATE.

XX Barwell JW, Longacre-Andre S, Mendis K, Nato F;

XX Roth C;

XX WPI; 1997-425033/39.

DR N-PSDB; AAT94549.

XX Recombinant protein containing the merozoite surface protein-1 p19 fragment - useful in anti-malarial vaccines, diagnosis and protein purification

XX Disclosure; Fig 1C; 85pp; French.

CC This is the amino acid sequence of a chimeric protein comprising amino acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1) linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.

CC The recombinant protein can be used for the production of anti-malarial vaccines, where the p19 fragment provides a high level of protective immunity since it includes epitopes not presented in the p42 fragment.

XX Sequence 127 AA;

Query Match 23.2%; Score 527; DB 18; Length 127;

Best Local Similarity 94.8%; Pred. No. 4.1e-30;

Matches 91; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 335 DMLNISQHCVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENGGCD 394

DB 32 DEFNISQHCVKKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENGGCD 91

QY 395 ADATCTEEDSGSRKKITCECTKPDSPYPLFDGIFCS 430

DB 92 ADAKTEEDSGSGNGKKTCECTKPDSPYPLFDGIFCS 127

RESULT 15

AAB37609

ID AAB37609 standard; Protein; 108 AA.

XX AAB37609;

XX 27-FEB-2001 (first entry)

XX Merozoite surface protein-119.

XX Merozoite surface protein; protozoacide; vaccine; malaria.

XX Plasmodium falciparum.

XX WO2000063245-A2.

XX 26-OCT-2000.

XX 20-APR-2000; 2000WO-GB01558.

XX 20-APR-1999; 99GB-0009072.

XX 13-MAY-1999; 99US-0311817.

XX 25-MAY-1999; 99CA-2271451.

XX (MEDI-) MEDICAL RES COUNCIL.

XX Holder A, Birdsell B, Feeney J, Morgan W, Syed S, Uthaipibull C;

XX WPI; 2001-015762/02.

XX N-PSDB; AAC68977.

XX Novel variants of the C-terminal fragment of Plasmodium merozoite surface protein-1, useful as vaccines for treating or preventing malaria -

XX Example 5; Fig 15; 126pp; English.

XX The present invention relates to non-natural variants of a C-terminal fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The non-natural variants have reduced affinity for at least 1 antibody capable of blocking a second antibody that inhibits the proteolytic

CC cleavage of Plasmodium MSP-1₄₋₂, and has the same affinity for at least
 CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
 CC MSP-1₄₋₂, compared to natural MSP-1₁₋₉. The non-natural variants of the
 CC present invention are useful for immunising a mammal against malaria, and
 CC can be used to treat malaria. The present sequence is MSP-119 protein.
 XX

SQ Sequence 108 AA;

Query Match 23.1%; Score 525; DB 22; Length 108;

Best Local Similarity 89.3%; Pred. No. 4.6e-30;

Matches 92; Conservative 4; Mismatches 3; Indels 4; Gaps 1;

Qy 329 IEGKQDMLNTSQHCVKQCPNSGCFRHLDERECKCLLYKQEGDKCVENPNTCNE 388

Db 8 IEGRH- ---NTAQHCVKQCPNSGCFRHLDERECKCLLYKQEGDKCVENPNTCNE 63

Qy 389 NNGGCDADATCTEEDSGSRKKTCTCKPDSYPLFDGIFCSS 431

Db 64 NNGGCDADAKCTEEDSGSGNGKKTCTCKPDSYPLFDGIFCSS 106

RESULT 16

AAB37608

ID AAB37608 standard; protein; 96 AA.

XX AAB37608;

XX 27-FEB-2001 (first entry)

XX Merozoite surface protein-1.

XX Merozoite surface protein; proteoacide; vaccine; malaria.

XX Plasmodium falciparum.

XX WO200063245-A2.

XX 26-OCT-2000.

XX 20-APR-2000; 2000WO-GB01558.

XX 20-APR-1999; 99GB-0009072.

XX 13-MAY-1999; 99US-0311817.

XX 25-MAY-1999; 99CA-2271451.

XX (MEDI-) MEDICAL RES COUNCIL.

XX Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaiipibull C;

XX WPT; 2001-015762/02.

XX Novel variants of the C-terminal fragment of Plasmodium merozoite

XX surface protein-1, useful as vaccines for treating or preventing

XX malaria -

XX Example 2; Page 48; 126pp; English.

XX The present invention relates to non-natural variants of a C-terminal

XX fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The

XX non-natural variants have reduced affinity for at least 1 antibody

XX capable of blocking a second antibody that inhibits the proteolytic

XX cleavage of Plasmodium MSP-1₄₋₂, and has the same affinity for at least

XX one third antibody that inhibits the proteolytic cleavage of Plasmodium

XX MSP-1₄₋₂, compared to natural MSP-1₁₋₉. The present sequence is the

XX wild-type MSP-1 protein. This sequence was used to generate the variants

XX of the present invention. The non-natural variants of the present

XX invention are useful for immunising a mammal against malaria, and can be

XX used to treat malaria.

XX Sequence 96 AA;

Query Match 23.0%; Score 524; DB 22; Length 96;

Best Local Similarity 95.7%; Pred. No. 4.7e-30;

Matches 90; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 338 NISOHCVKKQCPNSGCFRHLDERECKCLLYKQEGDKCVENPNTCENNGGCDADA 397

Db 1 NISOHCVKKQCPNSGCFRHLDERECKCLLYKQEGDKCVENPNTCENNGGCDADA 60

Qy 398 TCTEEDSGSRKKTCTCKPDSYPLFDGIFCSS 431

Db 61 KCTEEDSGSGNGKKTCTCKPDSYPLFDGIFCSS 94

RESULT 17

AAU00669

ID AAU00669 standard; Protein; 206 AA.

XX AAU00669;

XX 07-SEP-2001 (first entry)

XX Plasmodium vivax merozoite surface protein C-terminal region.

XX Merozoite surface protein; malaria; blood; serum; diagnosis; vaccine;

XX antigen-antibody composite; Enzyme Linked Immunosorbent Assay; ELISA;

XX MSP.

XX Plasmodium vivax.

XX WO200136587-A2.

XX 25-MAY-2001.

XX 15-NOV-2000; 2000WO-KR01302.

XX 15-NOV-1999; 99KR-0050616.

XX 25-APR-2000; 2000KR-0022041.

XX 20-MAY-2000; 2000KR-0027305.

XX (HUMA-) HUMANBIO CO LTD.

XX Park H;

XX WPT; 2001-343809/36.

XX N-PSDB; AAS00655.

XX New gene encoding merozoite surface protein of Plasmodium vivax, useful
 for producing protein for diagnosis of malaria and for vaccination -

XX Claim 5; Page 24; 24pp; English.

XX The sequence represents a Plasmodium vivax merozoite surface protein

XX (MSP) C-terminal region. The C-terminal region of the merozoite surface

XX protein has a strong antigenicity in malarial diseases. For diagnosis of

XX malaria, recombinant proteins with enhanced antigenicity, obtained by

XX addition of fusion proteins to surface protein C-terminal regions, can be

XX reacted with serum or blood of a Plasmodium infected patient

XX Antigen-antibody composites will be formed, and these are detected by

XX Enzyme Linked Immunosorbent Assay (ELISA). The recombinant antigens

XX provide a quick and reliable diagnosis of malaria, with good sensitivity

XX and selectivity.

XX Sequence 206 AA;

Query Match 17.8%; Score 406; DB 22; Length 206;

Best Local Similarity 41.0%; Pred. No. 3.4e-21;

Matches 80; Conservative 40; Mismatches 63; Indels 12; Gaps 4;

Qy 238 YNNLVNKIDDDYLINKAKINDCNVEKDEAHVKITKLSLKAIDDKIDLKKNPYDFEAIKK 297

Db 2 YESLVSKVNTYTDNLKKVINNCQLEKKEAITVKKLDQYNKMKDEKLEEK-----KSEKK 56

Qy 298 LINDDTKKDMLGKLTGLV--QNEPNTIISKIEGKQDMLNLSHQHCVKQCPNSGCF 356

Db 57 --NEVKSGLLEKLMKSLIKENESKEITLSOLLNVQTLTMSSEHTCIDTNVPDNRACY 114

KW Multivalent protein; immune response; Plasmodium vivax; parasite;
KW protozoacide; vaccine; malaria; recombinant; ViVac2.
OS Synthetic.
OS Plasmodium vivax.
XX WO200155181-A2.
XX 02-AUG-2001.
XX 29-JAN-2001; 2001WO-US02937.
XX 31-JAN-2000; 2000US-0179213.
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX Lal AA, Xiao L, Zhou Z;
XX WPI; 2001-514557/56.
DR N-PSDB; AAH47055.
XX New recombinant multivalent protein comprising antigenic determinants
PT derived from more than one stage in a life cycle of Plasmodium vivax,
PT useful as a vaccine for treating, preventing and reducing malarial
PT infection
XX Claim 5; Page 48-55; 59pp; English.
XX The invention relates to recombinant multivalent proteins (I) that
CC stimulate an immune response to Plasmodium vivax. (I) comprises antigenic
CC determinants, fragments or conservative substitutions, derived from more
CC than one stage in a life cycle of a Plasmodium vivax parasite. (I) is
CC useful as a vaccine for stimulating an immune response, specifically a
CC protective immune response that confers increased resistance to infection
CC by Plasmodium parasites, such as P. vivax. (I) is especially useful in
CC the treatment, prevention and reduction of malarial infection, as
CC research or diagnostic reagents for the detection of Plasmodium species
CC in a biological sample, and for conferring immunity against multiple
CC stages of the malarial parasite. The antibodies produced are useful for
CC the detection or measurement of antigenic epitopes derived from one or
CC more stages in a life cycle of a parasite, particularly P. vivax. The
CC vaccine comprising the recombinant proteins, is cost-effective, health-
CC promoting intervention for controlling, preventing or treating the
CC incidence of malaria. The present sequence represents the amino acid
CC sequence of the recombinant protein ViVac2p, a multivalent and
CC multistage vaccine against P. vivax.
XX
SQ Sequence 2028 AA;
Query Match 12.5%; Score 283.5; DB 22; Length 2028;
Best Local Similarity 24.2%; Pred. No. 3.9e-11;
Matches 86; Conservative 52; Mismatches 108; Indels 109; Gaps 11;
QY 88 SLKKQIEKNITFNILNLDILNSRLKRYFLDVLSDLMQFKHISSEYIIEISFKLLN 147
Db 1771 SVKRLKGN-FWICKLNAVY-----IEPQIVRWIREWRDYYSE----LPT 1813
QY 148 SEQNTLLSKYKIKESVENDIKF-----AQEGISYIEKVLAKYKD--DLESTKK 195
Db 1814 EVQK-----LKEKCDGKINTDKKVKVPCQACKSYDOWITRKNQWDVLSNKF 1864
QY 196 VIKKEKEFPSPPTPPSPAKTDEQKESKFLPFLTNIETLYNNLVNKIDDYILINKAK 255
Db 1865 ISVNAEKVQTAGIVTPYDILKQELDEFN-----EVAFENEINKRDGAVIELCV- 1913
QY 256 INDCNVEKDEAHVITKLSLDLKAIDDKIDLPKNYPDEAKIKLLINDTTKKDMLKLLSTG 315.
Db 1914 ---CSVEE-----AKNTQEVVTNVDNAKSPWTMS----- 1941
QY 316 LVQNPNTIISKIEGKFDQMLNLSOHCVKKQCPNSGCFRHLDERECKLLNYKQEG 375
Db 1942 -----SEHTCIDTNVDPNACRYILDGTEWRCLLTFRKEG 1977

QY 376 DKCVENPNTCENNGCGDADATCTEEDSGSRKKITCECTKPDSPFLFDGIFCS 430
Db 1978 GKCVPASNVTKDNNGGCAPEACKMTDS---NKIVCKTKEGSEPPFEGVCS 2028
RESULT 24
AA41355
ID AAR41355 standard; peptide; 49 AA.
XX AC AAR41355;
XX 04-MAR-1994 (first entry)
XX MSPLEGFIB EGF1-like domain variant.
XX Epidermal growth factor 1; merozoite surface protein 1; malaria;
KW vaccine.
XX Plasmodium yoelii.
XX Key Location/Qualifiers
FT Cleavage-site 1
FT /note= "introduced to facilitate cleavage
FT from recombinant protein"
XX WO9317107-A.
XX 02-SEP-1993.
XX 22-FEB-1993; 93WO-GB00367.
XX 22-FEB-1992; 92GB-0003821.
XX (MEDI-) MEDICAL RES COUNCIL.
XX Blackman MJ, Chappel JA, Holder AA;
XX WPI; 1993-288413/36.
XX Allelic variants of epidermal growth factor 1- or 2-like domains - of
PT merozoite surface protein 1, produced recombinantly for malaria
PT vaccines
XX Claim 1; Fig 1b; 35pp; English.
XX The sequence is that of an allelic variant of a merozoite surface
CC protein-1 epidermal growth factor (EGF) 1 like domain. It may be
CC used alone or as part of a fusion protein of EGF-1-like and
CC EGF-2-like domains in vaccines against malaria. When expressed
CC recombinantly it is produced in a form indistinguishable from
CC that in the native protein.
XX
SQ Sequence 49 AA;
Query Match 12.4%; Score 282; DB 14; Length 49;
Best Local Similarity 98.0%; Pred. No. 3.5e-13;
Matches 48; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 337 LNTISOHCVKKQCPNSGCFRHLDERECKLLNYKQEGDKCVENPNT 385
Db 1 MNISOHCVKKQCPNSGCFRHLDERECKLLNYKQEGDKCVENPNT 49
RESULT 25
AAB47487
ID AAB47487 standard; peptide; 108 AA.
XX AC AAB47487;
XX 15-NOV-2001 (first entry)
XX PV200C polypeptide.
XX

KW PV200C; merozoite surface protein; Plasmodium vivax; malaria;
 KW malaria-specific antibody; blood; antigen; protozoa; immunoassay.
 OS Plasmodium vivax.
 XX WO200161032-A1.
 PN 23-AUG-2001.
 XX 15-FEB-2001; 2001WO-KR00229.
 XX 17-FEB-2000; 2000KR-0007648.
 PR 17-FEB-2000; 2000KR-0007649.
 PR 17-FEB-2000; 2000KR-0007650.
 PR 10-MAR-2000; 2000KR-0012172.
 PR 08-AUG-2000; 2000KR-0045806.
 XX (GLDS) LG CHEM LTD.
 PA Lim K, Shon M, Yoo S, Lee S, Oh J, Lee S, Kim H;
 PI WPI; 2001-565352/63.
 DR Detecting malaria-specific antibodies in blood using antigens of
 PT malarial protozoa, useful for diagnosing malaria with long latent
 PT periods and low blood protozoa levels -
 XX Claim 11; Page 86-87; 89pp; English.
 PS This sequence is the polypeptide PV200C which is derived from the
 XX C-terminal portion of the merozoite surface protein of P. vivax. This
 CC polypeptide may be used in an immunoassay for malaria, which involves
 CC detecting malaria-specific antibodies in blood by using antigens of
 CC the malarial protozoa. The immunoassay has high specificity/sensitivity
 CC and may be used to diagnose types of malaria in which the latent
 CC period is long and the number of protozoa found in the blood is low.
 CC PV200C has high specificity to the antibodies and high purity. The
 CC PV200C surface protein has low pseudo-positive signals.
 XX Sequence 108 AA;
 SQ Query Match 12.3%; Score 280; DB 22; Length 108;
 Best Local Similarity 45.5%; Pred. No. 1.4e-12;
 Matches 51; Conservative 19; Mismatches 38; Indels 4; Gaps 1;
 QY 319 NFPTNTIISKIEGKFQDMLNISQHCVKQCPENSGCFRHLDERECKLLNYKQEGDKC 378
 Db 1 NESKEILSOLLNVQTLTWSSHTCIDTNVPDNCACYRDLDTETWRCLLTFKEGGKC 60
 QY 379 VNPNTCENNGGCDADATCTEEDSGSRKKITCTCTKPDSPYLPDGFICS 430
 Db 61 VPASNVTKDNNNGGCAPEACKMTDS----NKIVCKCTKEGSEPLFEGVFC 108
 RESULT 26
 AAR41354
 ID AAR41354 standard; peptide; 49 AA.
 XX AAR41354;
 AC AAR41354;
 XX 04-MAR-1994 (first entry)
 DT MSP1EGF1A EGF1-like domain variant.
 DE Epidermal growth factor 1; merozoite surface protein 1; malaria;
 XX vaccine.
 KW Plasmodium yoelii.
 OS Key Location/Qualifiers
 XX Cleavage-site 1 /note= "introduced to facilitate cleavage
 FT from recombinant protein"

XX WO9317107-A.
 PN 02-SEP-1993.
 PD 22-FEB-1993; 93WO-GB00367.
 PF 22-FEB-1992; 92GB-0003821.
 PR (MEDI-) MEDICAL RES COUNCIL.
 XX Blackman MJ, Chappel JA, Holder AA;
 PI WPI; 1993-288413/36.
 DR Allelic variants of epidermal growth factor 1- or 2-like domains - of
 PT merozoite surface protein 1, produced recombinantly for malaria
 PT vaccines
 XX Claim 1; Fig 1a; 35pp; English.
 PS The sequence is that of an allelic variant of a merozoite surface
 XX protein-1 epidermal growth factor (EGF) 1 like domain. It may be
 CC used alone or as part of a fusion protein of EGF-1-like and
 CC EGF-2-like domains in vaccines against malaria. When expressed
 CC recombinantly it is produced in a form indistinguishable from
 CC that in the native protein.
 XX Sequence 49 AA;
 SQ Query Match 12.3%; Score 279; DB 14; Length 49;
 Best Local Similarity 95.9%; Pred. No. 5.8e-13;
 Matches 47; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 337 LNTSQHCVKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNPT 385
 Db 1 MNISQHCVKQCPQNSGCFRHLDERECKLLNYKQEGDKCVENPNPT 49
 RESULT 27
 AAEL3033
 ID AAEL3033 standard; Protein; 289 AA.
 XX AAEL3033;
 AC AAEL3033;
 XX 28-JAN-2002 (first entry)
 DT Helicobacter pylori His-HOP38(-11) polypeptide.
 DE Polypeptide delivery system; immune stimulating complex; ISCOM;
 KW cholesterol; saponin; phospholipid; medicament; vaccine; therapy;
 KW Helicobacter pylori infection; HOP38; anti-bacterial.
 XX Helicobacter pylori.
 OS WO200176625-A1.
 XX 18-OCT-2001.
 PD 09-APR-2001; 2001WO-SE00799.
 PF 12-APR-2000; 2000GB-0008879.
 PR (ASTR) ASTRAZENECA AB.
 XX Shaprio A, Sanyal G;
 PI WPI; 2001-663016/76.
 DR N-PSDB; AAD21371.
 XX Producing a polypeptide delivery system useful in a vaccine to treat
 PT infection by mixing together the polypeptide, cholesterol, saponin, and
 PT a phospholipid in presence of a nonionic detergent and a second

PT detergent
 PS Disclosure: Page 40; 43pp; English.
 XX
 CC The invention relates to producing a polypeptide delivery system
 CC comprising an immune stimulating complex (ISCOM) coupled to a
 CC polypeptide. The method involves mixing the polypeptide, cholesterol,
 CC a saponin and a phospholipid in the presence of a nonionic detergent
 CC and a second detergent to form a solution and removing the detergent
 CC from the mixture to form the ISCOM. The polypeptide delivery system
 CC is used as a medicament for the manufacture of a vaccine for
 CC administration to mammalian patients, to treat and prevent
 CC Helicobacter pylori infection. The present sequence is His-HOP38(-11),
 CC a Helicobacter pylori polypeptide used in the polypeptide delivery
 CC system.
 XX
 SQ Sequence 289 AA;
 Query Match 12.2%; Score 278.5; DB 22; Length 289;
 Best Local Similarity 96.4%; Pred. No. 6.6e-12;
 Matches 53; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 OY 1 MHHHHHSSGLVPRGSGMKETAARFERQHMDSPDLGTDGDDKAMADIGSIEGRG 55
 DB 1 MHHHHHSSGLVPRGSGMKETAARFERQHMDSPDLGTDGDDKAMADIGS-EGDG 54
 RESULT 28
 AAB47732
 ID AAB47732 standard; Protein: 289 AA.
 XX
 AC AAB47732;
 XX
 DT 30-JAN-2002 (first entry)
 XX
 DE HOP38 #5.
 XX
 KW H. pylori; HOP38; polypeptide delivery system; polar head group;
 KW immune stimulating complex; ISCOM; antigenic fragment; saponin;
 KW cholesterol; phospholipid; detergent.
 OS Helicobacter pylori.
 XX
 PN WO200176623-A1.
 XX
 PD 18-OCT-2001.
 XX
 PF 09-APR-2001; 2001WO-SE00800.
 XX
 PR 12-APR-2000; 2000GB-0008877.
 XX
 PA (ASTR) ASTRAZENECA AB..
 XX
 PI Shapiro A, Sanyal G;
 XX
 DR WPI: 2002-025884/03.
 DR N-PSDB; AAH43782.
 XX
 CC Production of a polypeptide delivery system useful as a medicament
 CC comprises mixing together the polypeptide, cholesterol, saponin, and a
 CC phospholipid with a polar head group, in the presence of a detergent -
 PS Disclosure: Page 45; 48pp; English.
 XX
 CC The sequences given in AAB47728-32 are H. pylori HOP38 proteins
 CC which were used in the method of the invention. The method of the
 CC invention is a process for production of a polypeptide delivery system
 CC comprising an immune stimulating complex (ISCOM) coupled to a
 CC polypeptide of H. pylori or its antigenic fragment. The method comprises:
 CC mixing the polypeptide, cholesterol, a saponin, and a phospholipid
 CC having a polar head group; and removing the detergent from the mixture
 CC to form an ISCOM. The method has a broad applicability to polypeptides,
 CC including polypeptides that are unsuited to prior art processes.

XX
 SQ Sequence 289 AA;
 Query Match 12.2%; Score 278.5; DB 23; Length 289;
 Best Local Similarity 96.4%; Pred. No. 6.6e-12;
 Matches 53; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 OY 1 MHHHHHSSGLVPRGSGMKETAARFERQHMDSPDLGTDGDDKAMADIGSIEGRG 55
 DB 1 MHHHHHSSGLVPRGSGMKETAARFERQHMDSPDLGTDGDDKAMADIGS-EGDG 54
 RESULT 29
 AAM48247
 ID AAM48247 standard; Protein: 824 AA.
 XX
 AC AAM48247;
 XX
 DT 21-MAR-2002 (first entry)
 XX
 DE Thioredoxin/deoxyxylulose 5-phosphate synthase fusion protein.
 XX
 KW DXPS; deoxyxylulose 5-phosphate synthase; pyruvate; thioredoxin;
 KW glyceraldehyde 3-phosphate; plant growth modulator;
 KW microbial growth modulator; enzyme.
 XX
 OS Chimeric - Arabidopsis thaliana.
 OS Chimeric - Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Region 1..165
 FT /note= "Thioredoxin protein"
 FT Region 166..824
 FT /note= "Truncated DXPS protein"
 XX
 PN US6326164-B1.
 XX
 PD 04-DEC-2001.
 XX
 PF 27-JUL-2000; 2000US-0626589.
 XX
 PR 27-JUL-2000; 2000US-0626589.
 XX
 PA (PARA-) PARADIGM GENETICS INC.
 XX
 PI Rice JW, Kloti AS, Crawford JM, Lanning B, Stewart SJ;
 XX
 DR WPI: 2002-121106/16.
 DR N-PSDB; ABA95618.
 XX
 PT Assay for determining deoxyxylulose 5-phosphate synthase activity by
 PT measuring pyruvate depletion, useful for screening compounds that
 PT inhibit or enhance this activity which is useful for modulating plant
 PT and microbial growth -
 XX
 PS Claim 1; Columns 19-24; 26pp; English.
 CC
 CC The present sequence is a fusion protein comprising a truncated
 CC deoxyxylulose 5-phosphate synthase (DXPS) from Arabidopsis thaliana and
 CC thioredoxin (trxA). The N-terminal 58 amino acids of the DXPS protein
 CC were removed to generate the truncated protein. The truncated DXPS protein
 CC possesses DXPS activity. The full-length DXPS is given in AAM48245. The
 CC present invention relates to a method for determining DXPS-activity. The
 CC method comprises contacting pyruvate, and optionally glyceraldehyde
 CC 3-phosphate with DXPS and then determining the concentration of remaining
 CC pyruvate and/or glyceraldehyde 3-phosphate. The method is useful for
 CC screening for inhibitors and enhancers of DXPS activity which will have
 CC use as modulators of plant and microbial growth.

XX
 SQ Sequence 824 AA;
 Query Match 12.2%; Score 278.5; DB 23; Length 824;
 Best Local Similarity 37.4%; Pred. No. 2.7e-11;

AAY85150
 ID AAY85150 standard; protein; 467 AA.
 XX
 AC AAY85150;
 XX
 DT 23-JUN-2000 (first entry)
 XX
 DE Mouse secreted protein acidic and rich in cysteine (SPARC).
 XX
 KW SPARC; secreted protein acidic and rich in cysteine; pharmaceutical;
 KW drug; research reagent; mouse.
 XX
 OS Mus musculus.
 XX
 PN JP3012931-B1.
 XX
 PD 28-FEB-2000.
 XX
 PF 26-FEB-1999; 99JP-0049826.
 XX
 PR 26-FEB-1999; 99JP-0049826.
 XX
 PA (AGEN) KOGYO GIJUTSUINCHO.
 XX
 DR WPI; 2000-259560/23.
 DR N-PSDB; AAA09770.
 XX
 PT New secretion protein acidic and rich in cysteine (SPARC) for drugs and
 PT research reagents consists of specific amino acid sequence -
 XX
 PS Claim 1; Fig 1; 17pp; Japanese.
 XX
 CC This sequence represents the amino acid sequence of a mouse protein,
 CC having the physiological activity of SPARC (secreted protein acidic and
 CC rich in cysteine). The protein is used for the preparation of drugs,
 CC research reagents and pharmaceuticals. The SPARC protein is prepared
 CC efficiently and inexpensively in large quantities. The protein can be
 CC used in pharmaceuticals as it does no lose its physiological activity.
 XX
 SQ Sequence 467 AA;
 Query Match 12.1%; Score 275; DB 21; Length 467;
 Best Local Similarity 98.0%; Pred. No. 2.2e-11;
 Matches 50; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHSSGLVPRGSGMKETAARFQHMDSPLDGLTDDDDKAMADIGSI 51
 DB 116 MHHHHSSGLVPRGSGMKETAARFQHMDSPLDGLTDDDDKAMADIGSM 166
 RESULT 33
 AAY79067
 ID AAY79067 standard; protein; 467 AA.
 XX
 AC AAY79067;
 XX
 DT 20-JUN-2000 (first entry)
 XX
 DE Murine secreted protein acidic and rich in cysteine amino acid sequence.
 XX
 KW Secreted protein acidic and rich in cysteine; SPARC; mouse; nerve cell;
 KW neurocyte cell adhesion; cell protrusion retraction; cell shrinkage;
 KW nervous system disease; epilepsy; arteriosclerosis; wound healing.
 XX
 OS Mus sp.
 XX
 PN JP3012930-B1.
 XX
 PD 28-FEB-2000.
 XX
 PF 26-FEB-1999; 99JP-0049708.
 XX
 PR 26-FEB-1999; 99JP-0049708.

XX (AGEN) KOGYO GIJUTSUINCHO.
 XX
 DR WPI; 2000-306484/27.
 DR N-PSDB; AAZ98759.
 XX
 PT Drug composition for suppressing neurocyte cell adhesion, generating
 PT cell migration and promoting shrinkage retraction of the cell contains
 PT specific amino acid sequence -
 XX
 PS Claim 1; Fig 4; 21pp; Japanese.
 XX
 CC This sequence represents the amino acid sequence of the murine protein
 CC referred to as SPARC (secreted protein acidic and rich in cysteine). The
 CC protein is used as the active ingredient in a drug composition for
 CC suppressing neurocyte cell adhesion, generating cell migration and
 CC promoting a nerve protrusion shrinkage reaction. The composition can be
 CC used in wound healing and also to treat diseases of the nervous system,
 CC arteriosclerosis, and epilepsy.
 XX
 SQ Sequence 467 AA;
 Query Match 12.1%; Score 275; DB 21; Length 467;
 Best Local Similarity 98.0%; Pred. No. 2.2e-11;
 Matches 50; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHSSGLVPRGSGMKETAARFQHMDSPLDGLTDDDDKAMADIGSI 51
 DB 116 MHHHHSSGLVPRGSGMKETAARFQHMDSPLDGLTDDDDKAMADIGSM 166
 RESULT 34
 AAB26110
 ID AAB26110 standard; protein; 76 AA.
 XX
 AC AAB26110;
 XX
 DT 30-JAN-2001 (first entry)
 XX
 DE H. contortus clone 65e vector pET30a cloning junction #2.
 XX
 KW Nematode; parasite; helminth; sheep; goat; stomach; vaccine.
 XX
 OS Synthetic.
 XX
 PN WO200056763-A1.
 XX
 PD 28-SEP-2000.
 XX
 PF 16-MAR-2000; 2000WO-AU00210.
 XX
 PR 18-MAR-1999; 99AU-0009297.
 XX
 PA (NOVS) NOVARTIS AG.
 PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.
 XX
 PI Savin KW, Cook VR, Chen Y, Sexton JL, Apos E, Wilson LR;
 PI Griffiths TM, Newton SE;
 XX
 DR WPI; 2000-594573/56.
 DR N-PSDB; AAA94064.
 XX
 PT New Haemonchus contortus polypeptide for inducing a protective effect
 PT against a helminth by controlling helminth infection, growth, viability
 PT and/or egg fecundity and for ameliorating the symptoms of helminth
 PT infection -
 XX
 PS Disclosure; Fig 8; 94pp; English.
 XX
 CC The present sequence is the translated sequence found at the cloning
 CC junction in the pET30a vector of nematode Haemonchus contortus
 CC clone 65e. This clone encodes a novel protein. This organism is a
 CC parasite found in the stomach of its host (sheep and goats in

CC particular). The protein can be used in a vaccine against other
 CC helminths, as well as Haemonchus contortus, including trematodes,
 CC cestodes, nematodes and acanthocephala. These are all capable of causing
 CC severe illness in their hosts, which include sheep, pigs, goats, cattle,
 CC horses, donkeys, dogs, cats, guinea pigs and cage-birds, along with
 CC humans. Antibodies to the protein can be used to diagnose infection.
 CC Note: This sequence is stated as being the same as that shown in SEQ ID
 CC NO: 6 of the specification (see AAB26109). However, this sequence is
 CC shorter than the one shown here.

XX
 SQ Sequence 76 AA;

Query Match 12.0%; Score 274; DB 21; Length 76;
 Best Local Similarity 100.0%; Pred. No. 2.4e-12;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDGDDDKAMADIGS 50
 |||||
 Db 1 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDGDDDKAMADIGS 50

RESULT 35

AAB85253
 ID AAB85253 standard; Protein; 167 AA.

XX
 AC AAB85253;

XX
 DT 07-SEP-2001 (first entry)

XX Thioresoxin functional fragment.

XX Thioresoxin; porphobilinogen synthase; T-PPS; PPS; enzyme; herbicide;
 KW delta-aminolevulinic acid; plant growth inhibitor.

XX Lycopersicon esculentum.

XX WO200146446-A1.

XX 28-JUN-2001.

XX 19-DEC-2000; 2000WO-US34584.

XX 22-DEC-1999; 99US-0171785.

XX (PARA-) PARADIGM GENETICS INC.

XX Crawford JM, Rice J, Sevala V, Stewart S;

XX WPI; 2001-418081/44.

DR N-PSDB; AAH22802.

XX Novel plant thioresoxin-porphobilinogen synthase or porphobilinogen
 PT synthase polypeptides, useful for identifying compounds for use as
 PT herbicides by inhibiting enzymatic activity of the polypeptides -

XX Example 1; Page 19; 25pp; English.

XX The invention provides novel DNA sequences encoding enzymes such as plant
 CC thioresoxin-porphobilinogen synthase (T-PPS) and plant porphobilinogen
 CC synthase (PPS). Methods of determining the enzymatic activity of T-PPS or
 CC PPS or its functional fragment are provided that involves contacting
 CC delta-aminolevulinic acid with the protein or its functional fragment and
 CC measuring the amount of porphobilinogen formed from it. Compounds which
 CC can modify the enzymatic activity of T-PPS or PPS can also be identified
 CC similarly, which are useful for inhibiting plant growth by inhibiting
 CC enzymatic activity of T-PPS or PPS or its functional fragment. The
 CC compounds thus identified are useful as herbicides. The present sequence
 CC represents the thioresoxin functional fragment.

XX Sequence 167 AA;

Query Match 12.0%; Score 274; DB 22; Length 167;
 Best Local Similarity 100.0%; Pred. No. 6.7e-12;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDGDDDKAMADIGS 50
 |||||
 Db 116 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDGDDDKAMADIGS 165

RESULT 36

AAB85254
 ID AAB85254 standard; Protein; 167 AA.

XX
 AC AAB85254;

XX
 DT 07-SEP-2001 (first entry)

XX Thioresoxin functional fragment.

XX Thioresoxin; porphobilinogen synthase; T-PPS; PPS; enzyme; herbicide;
 KW delta-aminolevulinic acid; plant growth inhibitor.

XX Lycopersicon esculentum.

XX WO200146446-A1.

XX 28-JUN-2001.

XX 19-DEC-2000; 2000WO-US34584.

XX 22-DEC-1999; 99US-0171785.

XX (PARA-) PARADIGM GENETICS INC.

XX Crawford JM, Rice J, Sevala V, Stewart S;

XX WPI; 2001-418081/44.

DR N-PSDB; AAH22802.

XX Novel plant thioresoxin-porphobilinogen synthase or porphobilinogen
 PT synthase polypeptides, useful for identifying compounds for use as
 PT herbicides by inhibiting enzymatic activity of the polypeptides -

XX Example 1; Page 20; 25pp; English.

XX The invention provides novel DNA sequences encoding enzymes such as plant
 CC thioresoxin-porphobilinogen synthase (T-PPS) and plant porphobilinogen
 CC synthase (PPS). Methods of determining the enzymatic activity of T-PPS or
 CC PPS or its functional fragment are provided that involves contacting
 CC delta-aminolevulinic acid with the protein or its functional fragment and
 CC measuring the amount of porphobilinogen formed from it. Compounds which
 CC can modify the enzymatic activity of T-PPS or PPS can also be identified
 CC similarly, which are useful for inhibiting plant growth by inhibiting
 CC enzymatic activity of T-PPS or PPS or its functional fragment. The
 CC compounds thus identified are useful as herbicides. The present sequence
 CC represents the thioresoxin functional fragment.

XX Sequence 167 AA;

Query Match 12.0%; Score 274; DB 22; Length 167;

Best Local Similarity 100.0%; Pred. No. 6.7e-12;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDGDDDKAMADIGS 50
 |||||
 Db 116 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDGDDDKAMADIGS 165

RESULT 37

AAB85251
 ID AAB85251 standard; Protein; 551 AA.

XX
 AC AAB85251;

XX
 DT 07-SEP-2001 (first entry)

XX DE Plant thioedoxin-porphobilinogen synthase (T-PPS).
XX KW Thioedoxin; porphobilinogen synthase; T-PPS; enzyme; herbicide;
KW delta-aminolevulinic acid; plant growth inhibitor.
XX OS Lycopersicon esculentum.
XX XX
XX FH Key Location/Qualifiers
XX FT Protein 1..167
XX FT /note= "thioedoxin functional fragment"
XX FT Protein 168..551
XX FT /note= "pps"
XX XX
XX PN WO200146446-A1.
XX PD 28-JUN-2001.
XX PF 19-DEC-2000; 2000WO-US34584.
XX PR 22-DEC-1999; 99US-0171785.
XX PA (PARA-) PARADIGM GENETICS INC.
XX PI Crawford JM, Rice J, Sevala V, Stewart S;
XX DR WPI; 2001-418081/44.
XX DR N-PSDB; AAH22901.
XX XX
XX PT Novel plant thioedoxin-porphobilinogen synthase or porphobilinogen
XX PT synthase polypeptides, useful for identifying compounds for use as
XX PT herbicides by inhibiting enzymatic activity of the polypeptides -
XX XX
XX PS Disclosure; Page 16-17; 25pp; English.
XX CC The invention provides novel DNA sequences encoding enzymes such as plant
XX CC thioedoxin-porphobilinogen synthase (T-PPS) and plant porphobilinogen
XX CC synthase (PPS). Methods of determining the enzymatic activity of T-PPS or
XX CC PPS or its functional fragment are provided that involves contacting
XX CC delta-aminolevulinic acid with the protein or its functional fragment and
XX CC measuring the amount of porphobilinogen formed from it. Compounds which
XX CC can modify the enzymatic activity T-PPS or PPS can also be identified
XX CC similarly, which are useful for inhibiting plant growth by inhibiting
XX CC enzymatic activity of T-PPS or PPS or its functional fragment. The
XX CC compounds thus identified are useful as herbicides. The present sequence
XX CC represents the fusion protein T-PPS.
XX SQ Sequence 551 AA;
Query Match 12.0%; Score 274; DB 22; Length 551;
Best Local Similarity 100.0%; Pred. NO. 3.3e-11;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MHHHHHSSGLVPRGSGMKETAARFQHQHMDSPDLGTDGDDDDKAMADIGS 50
|||||
Db 116 MHHHHHSSGLVPRGSGMKETAARFQHQHMDSPDLGTDGDDDDKAMADIGS 165
RESULT 38
AAB85252
ID AAB85252 standard; Protein; 551 AA.
XX AAB85252;
XX XX
XX DT 07-SEP-2001 (first entry)
XX DE Plant thioedoxin-porphobilinogen synthase (T-PPS).
XX KW Thioedoxin; porphobilinogen synthase; T-PPS; enzyme; herbicide;
KW delta-aminolevulinic acid; plant growth inhibitor.
XX OS Lycopersicon esculentum.
XX XX

FH Key Location/Qualifiers
FT Protein 1..167
FT /note= "thioedoxin functional fragment"
FT Protein 168..551
FT /note= "pps"
XX XX
XX PN WO200146446-A1.
XX PD 28-JUN-2001.
XX PF 19-DEC-2000; 2000WO-US34584.
XX PR 22-DEC-1999; 99US-0171785.
XX PA (PARA-) PARADIGM GENETICS INC.
XX PI Crawford JM, Rice J, Sevala V, Stewart S;
XX DR WPI; 2001-418081/44.
XX DR N-PSDB; AAH22801.
XX XX
XX PT Novel plant thioedoxin-porphobilinogen synthase or porphobilinogen
XX PT synthase polypeptides, useful for identifying compounds for use as
XX PT herbicides by inhibiting enzymatic activity of the polypeptides -
XX XX
XX PS Claim 4; Page 17-18; 25pp; English.
XX CC The invention provides novel DNA sequences encoding enzymes such as plant
XX CC thioedoxin-porphobilinogen synthase (T-PPS) and plant porphobilinogen
XX CC synthase (PPS). Methods of determining the enzymatic activity of T-PPS or
XX CC PPS or its functional fragment are provided that involves contacting
XX CC delta-aminolevulinic acid with the protein or its functional fragment and
XX CC measuring the amount of porphobilinogen formed from it. Compounds which
XX CC can modify the enzymatic activity T-PPS or PPS can also be identified
XX CC similarly, which are useful for inhibiting plant growth by inhibiting
XX CC enzymatic activity of T-PPS or PPS or its functional fragment. The
XX CC compounds thus identified are useful as herbicides. The present sequence
XX CC represents the fusion protein T-PPS.
XX SQ Sequence 551 AA;
Query Match 12.0%; Score 274; DB 22; Length 551;
Best Local Similarity 100.0%; Pred. NO. 3.3e-11;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MHHHHHSSGLVPRGSGMKETAARFQHQHMDSPDLGTDGDDDDKAMADIGS 50
|||||
Db 116 MHHHHHSSGLVPRGSGMKETAARFQHQHMDSPDLGTDGDDDDKAMADIGS 165
RESULT 39
AAB37611
ID AAB37611 standard; protein; 93 AA.
XX AAB37611;
XX AC AAB37611;
XX DT 27-FEB-2001 (first entry)
XX DE Merozoite surface protein-1.
XX KW Merozoite surface protein; protazoacide; vaccine; malaria.
XX OS Plasmodium vivax.
XX PN WO200063245-A2.
XX PD 26-OCT-2000.
XX XX
XX PF 20-APR-2000; 2000WO-GB01558.
XX PR 20-APR-1999; 99GB-0009072.
XX PR 13-MAY-1999; 99US-0311817.
XX PR 25-MAY-1999; 99CA-2271451.

[illegible]

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 19, 2003, 14:17:05 ; Search time 218 Seconds
(without alignments)
4452.345 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 2275

Sequence: 1 MHHHHHHSLVPRGSMKE.....TCECTKPDSPYLDGIFCSS 431

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US10057531/runat_12052003.091030.21654/app_query.fasta_1.503
-DB=N.Geneseq.101002 -QFMT=fastp -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10057531 -RCGN_1_1_200_@runat_12052003.091030.21654 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.Geneseq.101002.*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1970	86.6	1128	22	DNA encoding major
2	1931.5	84.9	1181	16	Plasmodium falcipara
3	1175	51.6	1897	18	Chimeric MSA-1 ant
4	1175	51.6	1950	18	Chimeric MSA-1 ant
5	1175	51.6	1950	18	Chimeric MSA-1 ant
6	1175	51.6	1950	18	Chimeric MSA-1 ant
7	1161.5	51.1	1065	20	P. falciparum mod
8	1161.5	51.1	1065	20	Chimeric MSA-1 ant
9	1161.5	51.1	1065	20	Chimeric MSA-1 ant
10	1161.5	51.1	1065	20	Chimeric MSA-1 ant
11	1161.5	51.1	1065	20	Chimeric MSA-1 ant
12	1161.5	51.1	1065	20	Chimeric MSA-1 ant
13	1154.5	50.7	1140	20	Chimeric MSA-1 ant
14	1154.5	50.7	1140	20	Chimeric MSA-1 ant
15	1066	46.9	5760	6	Chimeric MSA-1 ant
16	640.5	28.2	786	22	Chimeric MSA-1 ant
17	527	23.2	354	18	Chimeric MSA-1 ant
18	527	23.2	354	18	Chimeric MSA-1 ant
19	527	23.2	354	18	Chimeric MSA-1 ant
20	527	23.2	354	18	Chimeric MSA-1 ant
21	525	23.1	333	22	Chimeric MSA-1 ant
22	521	22.9	309	24	Chimeric MSA-1 ant
23	521	22.9	309	24	Chimeric MSA-1 ant
24	521	22.9	309	24	Chimeric MSA-1 ant
25	500	22.0	3147	24	Chimeric MSA-1 ant
26	500	22.0	3147	24	Chimeric MSA-1 ant
27	500	22.0	3147	24	Chimeric MSA-1 ant
28	406	17.8	618	22	Chimeric MSA-1 ant
29	284.5	12.5	5438	22	Chimeric MSA-1 ant
30	284.5	12.5	5438	22	Chimeric MSA-1 ant
31	284.5	12.5	5438	22	Chimeric MSA-1 ant
32	284.5	12.5	5438	22	Chimeric MSA-1 ant
33	278.5	12.2	870	22	Chimeric MSA-1 ant
34	278.5	12.2	870	22	Chimeric MSA-1 ant
35	278.5	12.2	870	22	Chimeric MSA-1 ant
36	277.5	12.2	2079	21	Chimeric MSA-1 ant
37	275	12.1	7375	21	Chimeric MSA-1 ant
38	275	12.1	7375	21	Chimeric MSA-1 ant
39	274	12.0	241	21	Chimeric MSA-1 ant
40	274	12.0	428	22	Chimeric MSA-1 ant
41	274	12.0	501	22	Chimeric MSA-1 ant
42	274	12.0	589	22	Chimeric MSA-1 ant
43	274	12.0	1653	22	Chimeric MSA-1 ant
44	274	12.0	6553	24	Chimeric MSA-1 ant
45	274	12.0	6618	24	Chimeric MSA-1 ant

ALIGNMENTS

RESULT 1
AAF89840
ID AAF89840 standard; DNA; 1128 BP.
XX
AC AAF89840;
XX
DT 23-JUL-2001 (first entry)
XX
DE DNA encoding major merozoite surface protein-1 fragment of 42kDa.
XX
KW Major merozoite surface protein-1; MSP1-42; melittin signal peptide;
KW malaria vaccine; ss.
XX
OS Plasmodium falciparum.
XX
FH Key
FT CDS
FT 1..1128
FT /*tag= a

XX 29-JAN-1997; 97WO-US01395.
 XX 29-JAN-1996; 96US-0593006.
 XX (GEOU) UNIV GEORGETOWN.
 PA Davidson EA, Yang S;
 XX WPI; 1997-393372/36.
 XX Malaria vaccine - comprises expression vector expressing fragment of
 PT merozoite surface antigen
 XX Claim 33; Fig 4; 75pp; English.
 CC The present sequence encodes a chimeric protein that contains a
 CC human anchor sequence, and a merozoite surface antigen-1 (MSA-1) carboxy
 CC terminal peptide. The C-terminal fragment of MSA-1 provides a more
 CC specific response than the complete MSA-1, and attachment of the
 CC anchor sequence improve the immunogenicity of the protein better than
 CC the use of an adjuvant. The chimeric protein, and expression vectors
 CC (analogues that express MSA-1 without either signal or anchor peptides),
 CC particularly in the form of recombinant vaccinia virus, are used in
 CC vaccines to prevent or treat malaria caused by Plasmodium falciparum. The
 CC vaccinia vector expresses the antigen fragment for many days, or even
 CC years, generating a long-lasting immune response (humoral and/or
 CC cell-mediated) against the merozoite form of the parasite, in humans or
 CC other animals.
 XX SQ Sequence 1897 BP; 815 A; 230 C; 274 G; 578 T; 0 other;

Alignment Scores:

Pred. No.: 7,05e-89 Length: 1897
 Score: 1175.00 Matches: 226
 Percent Similarity: 71.43% Conservative: 59
 Best Local Similarity: 56.64% Mismatches: 90
 Query Match: 51.65% Indels: 24
 DB: 18 Gaps: 4

US-10-057-531a-2 (1-431) x AAT97958 (1-1897)

QY 34 ProAspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGly 53
 DB 706 CCTATATTGGAGATCCGAGAGAT-----TATGATGATTAGGACAGTAGTAACA 759
 QY 54 ArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGlu 73
 DB 760 GGAGAAGCAGTAACCTCTCCGTA---ATTGATAACATACCTTCTTAAATTTGAAATCAA 816
 QY 74 TyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIle 93
 DB 817 TATGAGGTTTATATTTAAACCTTTACAGGTGTTTATAGAAGTTTAAACAAATTA 876
 QY 94 GluLysAsnIlePheThrPheAsnLeuAsnLeuAspIleLeuAsnSerArgLeuLys 113
 DB 877 GAAATACGTTATGACATTTAAGTTAATGTTAAGGATATTTAAATTCAGATTAAAT 936
 QY 114 LysArgLysTyrPheLeuAspValLeuSerAspLeuMetGlnPheLysHisIleSer 133
 DB 937 AAACGTGAAATTTCAAAATGTTTACAAATCAGATTTAATTCATATAAGATTAAACA 996
 QY 134 SerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThr 153
 DB 997 TCAAGTATATGTTGTCAAGATCCATATAATTTCTTAATAAGAAAAAAGAGATAAA 1056
 QY 154 LeuLeuLysSerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGln 173
 DB 1057 TTCTAACGAGTTAATATATTAGGATTCATAGATACGGATATAAATTTGCAAAAT 1116
 QY 174 GluGlyIleSerTyrTyrGlyLysValLeuAlaLysTyrLysAspAspLeuGluSerIle 193
 DB 1117 GATGTTCTTGGATATATAAAATATATCCGAAAAATATAATCAGATTAGATTCAATT 1176

QY 194 LysLysValIleLysGluGluLysPheProSerSerProThrThrProPro 213
 DB 1177 AAAAAATATATC----- 1188
 QY 214 SerProAlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsn 233
 DB 1189 -----AACGCAAAACAAGGTGAAAATGAGAAATACCTTCCTTTTAAACAAT 1236
 QY 234 IleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLys 253
 DB 1237 ATTGAGACCTTATATAAACAGTTAATGATAAATTTGATTATTGTAATTCATTAGAA 1296
 QY 254 AlaLysIleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeu 273
 DB 1297 GCAAAAGTTCTAAATTTATACATATGAGAAATCAAAAGTAGAAGTTAAATAAAGAACTT 1356
 QY 274 SerAspLeuLysAlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGlu 293
 DB 1357 AATTACTTTAAACAAATTTCAAGACAAATTTGCAGATTTTAAACAAAATAACAATTTCTGT 1416
 QY 294 AlaIleLysLysLeuIleAsnAspThrLysLysAspMetLeuGlyLysLeuSer 313
 DB 1417 GGAATTGCTGATTATCAACAGATTATAACCATATACTATTACACAAAGTCTCTTACT 1476
 QY 314 ThrGlyLeuVal---GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLys 332
 DB 1477 ACAGTATGTTTGTGAAATCTTCTGCTAAACCGTTTATCTAATTTACTTGTGAGAAAC 1536
 QY 333 PheGlnAspMetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsn 352
 DB 1537 TTGCAAGGTATGTTTAAACATTTTCACACACCAATGCGTTAAACAAATATGTCACAAAT 1596
 QY 353 SerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCysLeuAsnTyrLys 372
 DB 1597 TCTGGATGTTTCACACATTTAGATGAAAGACAGAAAGTAATGTTATTATAATACAAA 1656
 QY 373 GlnGluGlyAspLysCysValGluAsnProAsnProThrCysAsnGluAsnGlyGly 392
 DB 1657 CAAGAAGGTGATAAATGTTGAAATCCAAATCCTACTTGTAAACGAAATAATGTTGGA 1716
 QY 393 CysAspAlaAspAlaThrCysThrGluGluAspSerGlySerArgLysIleThr 412
 DB 1717 TGTGATGAGATGCCAAATGTACCAAGAGATTACAGTAGCAACGGAAGAAATACACA 1776
 QY 413 CysGluCysThrLysProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
 DB 1777 TGTGAATGTACTAAACCTGATTCTTATCCACTTTTCGATGTTGTTCTGTCAGTTCC 1833
 RESULT 4
 AAT97956
 ID AAT97956 standard; DNA; 1950 BP.
 XX
 AC AAT97956;
 XX
 DT 03-APR-1998 (first entry)
 XX
 DE Chimeric MSA-1 antigenic protein 1 used in a malaria vaccine.
 XX
 KW Signal peptide; malaria; vaccine; merozoite surface antigen-1 peptide;
 KW MSA-1; recombinant vaccinia virus; Plasmodium falciparum; anchor peptide;
 KW immune response; humoral; cell-mediated; merozoite; ss.
 XX Chimeric - Mammalia.
 OS Chimeric - Plasmodium falciparum.
 OS Chimeric - Homo sapiens.
 XX Key Location/Qualifiers
 FT misc_feature 1..165 /tag= a
 FT /note= "signal sequence added to improve immunogenicity"
 FT misc_feature 166..1893


```

FT      /*tag= b
FT      /note= "Plasmodium falciparum MSA-1 peptide"
FT      1894..1950
FT      /*tag= c
FT      /note= "anchor sequence added to improve
XX      immunogenecity"
PN      W09726911-Al.
XX
XX      31-JUL-1997.
XX
XX      29-JAN-1997; 97WO-US01395.
XX
XX      29-JAN-1996; 96US-0593006.
XX
XX      (GEU ) UNIV GEORGETOWN.
XX
XX      Davidson EA, Yang S;
XX
XX      WPI; 1997-393372/36.
XX
XX      Malaria vaccine - comprises expression vector expressing fragment of
XX      merozoite surface antigen
XX
XX      Claim 33; Fig 2; 75pp; English.
XX
XX      The present sequence encodes a chimeric protein that contains a
XX      mammalian signal and an anchor sequence, and a merozoite surface
XX      antigen-1 (MSA-1) carboxy terminal peptide. The C-terminal fragment of
XX      MSA-1 provides a more specific response than the complete MSA-1,
XX      and attachment of anchor and signal sequences improve the immunogenicity
XX      of the protein better than the use of an adjuvant. The chimeric protein,
XX      and expression vectors (analogues that express MSA-1 without either
XX      signal or anchor peptides), particularly in the form of recombinant
XX      vaccinia virus, are used in vaccines to prevent or treat malaria caused
XX      by Plasmodium falciparum. The vaccinia vector expresses the antigen
XX      fragment for many days, or even years, generating a long-lasting immune
XX      response (humoral and/or cell-mediated) against the merozoite form of
XX      the parasite, in humans or other animals.
XX
XX      Sequence 1950 BP; 830 A; 236 C; 277 G; 607 T; 0 other;
SQ
Alignment Scores:
Pred. No.:      7.29e-89      Length:      1950
Score:          1175.00      Matches:      226
Percent Similarity: 71.43%      Conservative: 59
Best Local Similarity: 56.64%      Mismatches: 90
Query Match:      53.65%      Indels:      24
DB:              16      Gaps:         4
US-10-057-531A-2 (1-431) x AAT97956 (1-1950)
QY      34 ProAspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGly 53
DB      760 CCTATATTGGAGATCCGAGAGAT-----TATGATGATTTAGGCAAGATGATGAACA 813
QY      54 ArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGlu 73
DB      814 GCAGAGACGAGTAACCTCTCCGTA---ATTGATAACATCTTCTTAAATTTGAAATGAA 870
QY      74 TyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIle 93
DB      871 TATGAGGTTTATATTAAACCTTTAGCAGGTTTATAGAGTTTAAAGAAATTA 930
QY      94 GluLysAsnIlePheThrPheAsnLeuAsnLeuAspIleLeuAsnSerArgLeuLys 113
DB      931 GAAATAACGTTATGACATTTATGTTTATGTTAAGGATATTTAAATTCACGATTTAAT 990
QY      114 LysArgLysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSer 133
DB      991 AAACGTGAAATTTCAAAAATGTTTAGAATCAGATTTAATTCATATAAAGATTTAACA 1050
QY      134 SerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThr 153

```

```

DB      1051 TCAAGTAATATCTTCTCAAGATCCATATAAATTTCTTAATAAGAAAAAGAGATAAA 1110
QY      154 LeuLeuLysSerTyrIleLysGluSerValGluAsnAspIleLysPheAlaGln 173
DB      1111 TTCTTAAGCAGTTATAATTAATTAAGGATTCATACGATGATATAAATTTGCAAT 1170
QY      174 GluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIle 193
DB      1171 GATGTTCTTGGATATTATAAATATATATCCGAAAAATATAAATCAGATTAGATTCAAT 1230
QY      194 LysLysValIleLysGluLysGluLysPheProSerSerProProThrThrProPro 213
DB      1231 AAAAAATATATC----- 1242
QY      214 SerProAlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsn 233
DB      1243 -----AACGCAAAACAAAGGTGAAATGAGAAATACCTCCCTTTTAAACAAT 1290
QY      234 IleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLys 253
DB      1291 ATTGACACCTTATATAAACAGCTTAATGATAAATTTGATTATTGTAATTCATTTAGAA 1350
QY      254 AlaLysIleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeu 273
DB      1351 GCAAAAGTTCTAAATTTATACATATGAGAAATCAACGCTAGAACTTAAATAAAGAACTT 1410
QY      274 SerAspLeuLysAlaIleAspLysIleAspLeuPheLysAsnProTyrAspPheGlu 293
DB      1411 AATTACTTAAACAACTTCAAGACAAATTTGGCAGATTTTAAAAAAAATAACAAATTCGTT 1470
QY      294 AlaIleLysLysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuSer 313
DB      1471 GGAATTTGCTGATTTATCAACAGATTATACCATTAATACTTATGACAAAGTTCCTTAGT 1530
QY      314 ThrGlyLeuVal---GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLys 332
DB      1531 ACAGGTATGTTTGTGAAATCTTGCTAAACCGTTTATCTAATTTACTTGTATGGAAC 1590
QY      333 PheGlnAspMetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsn 352
DB      1591 TTGCAAGGTATGTTTAAACATTTTCAACACCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1650
QY      353 SerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLys 372
DB      1651 TCTGGATGTTTCAGACATTTAGATGAAGAGAGATGTAATGTTTATTAATTTACAAA 1710
QY      373 GlnGluGlyAspLysCysValGluAsnProAsnProThrCysAsnGluAsnGlyGly 392
DB      1711 CAAGAAGGTGATAAATGTTGTAATAATCCAAATCCCTACTTGTAAACGAAATAATGTTGA 1770
QY      393 CysAspAlaAspAlaThrCysThrGluAspSerGlySerSerArgLysLysIleThr 412
DB      1771 TGTGATGTCAGATGCCAAATGTACCGAAGAGATTTCAGGTAGCAACGAAAGAAATTCACA 1830
QY      413 CysGluCysThrLysProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
DB      1831 TGTGAATGTACTAAACCTGATTCTTATCCACTTTTCGATGGTATTTCCTGCACTTCC 1887
RESULT 5
AAV21451
ID      AAV21451 standard; DNA; 4940 BP.
XX
XX      AAV21451;
AC
XX      23-SEP-1998 (first entry)
DT
XX      P. falciparum modified gp190 DNA.
DE
XX      gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW      monoclonal antibody; passive immunisation; parasite; ss.
XX
XX      Plasmodium falciparum.
OS

```



```

XX OS Plasmodium falciparum.
XX OS Synthetic.
XX PN WO9920774-A2.
XX XX
XX PD 29-APR-1999.
XX XX
XX PF 20-OCT-1998; 98WO-US22226.
XX XX
XX PR 15-MAY-1998; 98US-0085649.
XX PR 20-OCT-1997; 98US-0062592.
XX PA (GENZ ) GENZYME TRANSGENICS CORP.
XX XX
XX PI Chen LH, Meade H;
XX XX
XX DR WPI; 1999-288313/24.
XX DR P-PSDB; AAY09372.
XX XX
XX PT Modified malarial protein for use in anti-malarial vaccines
XX XX
XX PS Claim 1; Fig 1; 35pp; English.
XX XX
XX CC This novel, modified nucleic acid encodes the 42 kDa C-terminal
XX CC portion (see AAY09372) of malaria merozoite surface protein MSP-1
XX CC (MSP-1-42), an important target for the development of a vaccine
XX CC against Plasmodium falciparum. The nucleic acid sequence has been
XX CC modified compared to the native sequence of MSP-1-42 (see AAX56009)
XX CC such that 306 nucleotide positions have been replaced to lower the
XX CC AT content (from 76 to 49.7%) and to eliminate 10 mRNA instability
XX CC motifs while maintaining the same protein amino acid sequence.
XX CC These alterations allow MSP-1-42 to be expressed in mammalian cell
XX CC culture and in transgenic mice. Native MSP-1-12 is known to be
XX CC difficult to express in cell culture systems, mammalian cell
XX CC culture systems or in transgenic animals. The invention allows
XX CC expression of MSP-1-42 protein in the milk of transgenic animals,
XX CC and also provides a DNA vaccine comprising a vector containing the
XX CC altered MSP-1-42 sequence.
XX SQ Sequence 1065 BP; 324 A; 254 C; 280 G; 207 T; 0 other;

Alignment Scores:
Pred. NO.: 4,71e-88 Length: 1065
Score: 1161.50 Matches: 218
Percent Similarity: 73.71% Conservative: 54
Best Local Similarity: 59.08% Mismatches: 76
Query Match: 51.05% Indels: 21
DB: 20 Gaps: 2

US-10-057-531a-2 (1-431) x AAX56008 (1-1065)
Oy 63 MetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeu 82
Db 19 ATCGATAACATCTCTCCCAAGATCGAAGACGAGTACGAGGTGCTGTACCTGAAGCGCGTG 78
Oy 83 AlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeu 102
Db 79 GCAGGGGTCTACCGGAGCGCTGAAGAAGACGCTGGAGAACACACGTGATGACCTTCAACGTG 138
Oy 103 AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeu 122
Db 139 AACGTGAAGGATATCTCTGAACAGCGCGTTCACACAGCGGAGCACTTCAAGACGTGCTG 198
Oy 123 GluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSer 142
Db 199 GAGACGATCTGATCCCTACCAAGGATCTGACCAACGACCACTAGTGTGTCAGGATCC 258
Oy 143 PheLysLeuLeuAsnSerGlnLysAsnThrLeuLysSerTyrLysTyrIleLys 162
Db 259 TACAGTTCTTCAACAGGAAGAGATAGTTCTCTGAGCAGTACCAACTACATCAAG 318
Oy 163 GluSerValGluAsnAspIleLysPheAlaGlnGluLysSerTyrTyrGluLysVal 182

```

```

Db 319 GATAGCATTTGATACCGGATATCAACTTCGCCAACAGATGTCTCTGGGATACTACAAGATCCTG 378
Oy 183 LeuAlaLysTyrLysAspLeuGluSerIleLysValIleLysValIleLysValIleLysVal 202
Db 379 TCCGAGAAAGTACAAGAGCGATCTGGATTCAATCAAGAAAGTACATC----- 423
Pd 203 LysPheProSerSerProThrThrProProSerProAlaLysThrAspGluGlnLys 222
Db 424 -----AACGATAAGCAGGGA 438
Oy 223 LysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuVal 242
Db 439 GAGAACGAGAGTACTCTGCTCTCTGATCACCTGGAGGCCAAGGTCTCTGAACTACATATGAG 498
Oy 243 AsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGlu 262
Db 499 GATAAGATTGATCTGCTGATCACCTGGAGGCCAAGGTCTCTGAACTACATATGAG 558
Oy 263 LysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaLysAspLys 282
Db 559 AAGACCAACGTGGAGTCAAGATCAAGAGCTGAATTACCTGAAGACCATCCAGGATAAG 618
Oy 283 IleAspLeuLysAsnProTyrAspPheGluAlaIleLysLysLysLysLysLysLys 302
Db 619 CTGGCCGATTTCAAGAACAAACAACTTCGTCCGGGATCGCCGATCTGAGCACCGATTAC 678
Oy 303 ThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal---GlnAsnPhePro 321
Db 679 AACCAACAACCTCTCTGACCAAGTTCCTGAGCACCGGTATGCTTCCAAAACCTGGCC 738
Oy 322 AsnThrIleLeuSerLysLysLeuIleGluLysPheGlnAspMetLeuAsnIleSerGln 341
Db 739 AAGACCGTCTGAGCAACCTGCTGGATGGGAACCTGCGAGGGGATGCTGAACATCAGCCAG 798
Oy 342 HisGlnCysValLysLysLysCysProGluAsnSerGlyCysPheArgHisLeuAspGlu 361
Db 799 CACCAAGTGTGAAGAACGAGTGTCCCAAGACACGGGTGTTTCAGACACCTGGATGAG 858
Oy 362 ArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluLysLysLysCysValGluAsn 381
Db 859 AGAGAGGAGTGTAAAGTGTCTGTAACCTACAAGCAGGAAGGTGATAAGTGTGCAAAAC 918
Oy 382 ProAsnProThrCysAsnGluAsnAsnGlyLysCysAspAlaAspAlaThrCysThrGlu 401
Db 919 CCAATCTCTACTTGTACGAGAACAAATGTTGGATGTGATGCGGATGCCAAGGTGATCCGAG 978
Oy 402 GluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLysProAspSerTyr 421
Db 979 GAGGATTCAGGAGGACGCGGAGAGAGATCACCTGTGAGTGTACCAAGCCTGATCTTAT 1038
Oy 422 ProLeuPheAspGlyIlePheCysSer 430
Db 1039 CCACGTGTCGATGGTATCTTCTGTAGT 1065

RESULT 8
AAX25586
ID AAX25586 standard; cDNA; 1065 BP.
XX
XX AAX25586;
XX
XX 02-AUG-1999 (first entry)
XX
XX Merozoite surface protein MSP-1-42 modified cDNA.
XX
XX MSP-1: merozoite surface protein; malaria; vaccine;
XX protein engineering; protein expression; codon usage;
XX transgenic animal; ss.
XX
XX Plasmodium falciparum.
XX OS Synthetic.
XX PN WO9920766-A2.

```


PR 20-OCT-1997; 97US-0062592.
 XX (GENZ) GENZYME TRANSGENICS CORP.
 XX Chen LH, Meade H;
 XX WPI; 1999-288313/24.
 DR P-PSDB; AAY09373.
 XX
 PT Modified malarial protein for use in anti-malarial vaccines
 XX
 PS Example; Fig 2; 35pp; English.
 XX
 CC This nucleic acid encodes a 42 kDa C-terminal portion (see AAY09373)
 CC of malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
 CC important target for the development of a vaccine against
 CC Plasmodium falciparum. The 3' end of the sequence has been
 CC modified to include a 6xHis tag. The nucleic acid sequence has
 CC been modified (see AX56008) according to a method of the invention
 CC in order to improve expression in mammalian host cells and in
 CC transgenic animals. In the modified coding sequence, 306 nucleotide
 CC positions have been replaced to lower the AT content (from 76 to
 CC 49.7%) and to eliminate 10 mRNA instability motifs. The encoded
 CC amino acid sequence is unaltered. In another modified sequence
 CC (see AX56021), a signal peptide sequence has been added and two
 CC N-glycosylation sites eliminated. The invention allows expression
 CC of MSP-1-42 protein in the milk of transgenic animals, and also
 CC provides a DNA vaccine comprising a vector containing the altered
 CC MSP-1-42 sequence.
 XX
 SQ Sequence 1088 BP; 454 A; 139 C; 150 G; 345 T; 0 other;
 Alignment Scores:
 Pred. No.: 4.83e-88 Length: 1088
 Score: 1161.50 Matches: 218
 Percent Similarity: 73.71% Conservative: 54
 Best Local Similarity: 59.08% Mismatches: 76
 Query Match: 51.05% Indels: 21
 DB: 20 Gaps: 2
 US-10-057-531A-2 (1-431) x AAX56009 (1-1088)
 QY 63 MetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeu 82
 DB 19 ATTGATAACACTACTTCTCTAAATTTGAAATGAATGACGAGGTTTATATTATAAACCTTTA 78
 QY 83 AlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeu 102
 DB 79 GCAGGTGTTATAGAGTTTAAATAAACCAATAGAAATAACGTTATGACATTTAATGTT 138
 QY 103 AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeu 122
 DB 139 AATGTTAAGGATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 198
 QY 123 GluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSer 142
 DB 199 GAATCAGATTTAATTCATATAAGATTTAATCAATCAAGTAATATGTTCTCAAGATCCA 258
 QY 143 PheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLys 162
 DB 259 TATAAATTTCTTAATAAGAAAGAAAGATATAATTTCTTAAGCAGTTATTAATATAATAG 318
 QY 163 GluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLysVal 182
 DB 319 GATTCATAGATAGCATATAATTTTGCAATGATGTTCTTGGATATTATAAANTATTA 378
 QY 183 LeuAlaLysTyrLysAspLeuGluSerIleLysValIleLysGluLysGlu 202
 DB 379 TCCGAAAAATATAAATCAGATTAGATTCAATTAATAAATAATATC----- 423
 QY 203 LysPheProSerSerProThrThrProProSerProAlaLysThrAspGluGlnLys 222
 DB 424 -----AACGACAAACAAGGT 438

QY 223 LysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuVal 242
 DB 439 GAAATGAGAAATACCTTCCCTTTTAAACAATATTGAGACCTTATATAAACAGTTAAT 498
 QY 243 AsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGlu 262
 DB 499 GATAAAATGATTTATTGTAATTCATTAGAACCAAAAGTTCTTAATATATACATATGAG 558
 QY 263 LysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAspAspLys 282
 DB 559 AAATCAACGTCAGAGTTAAATAAAGAACTTAATTTACTTAAAAACAATTTCAAGCAAA 618
 QY 283 IleAspLysAspAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAspAsp 302
 DB 619 TTGGCAGATTTTAAAAAATAACAAATTTCTGTTGGAATTCCTGATTTATCAACAGATTAT 678
 QY 303 ThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal---GlnAsnPhePro 321
 DB 679 AACCATTAATTAACATTATTGACAAAGTTCTTAGTACAGGTATGGTTTTTGAATAATCTT 738
 QY 322 AsnThrIleIleSerLysLeuIleGluLysPheGlnAspMetLeuAsnIleSerGln 341
 DB 739 AAACCGTTTATCTAATTTACTTTGATGGAACCTTCAGGTATGTTAAACATTTCAACA 798
 QY 342 HisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGlu 361
 DB 799 CACCAATGCTAAAAAACAATGTCACAAAAATTTCTGGATGTTTTCAGACATTTAGATGAA 858
 QY 362 ArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluLysLysCysValGluAsn 381
 DB 859 AGAAGAAGATGTAATGTTTATTTAAATTTACAAACAAGAGGTGATAAATGTTGTTGAAAT 918
 QY 382 ProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGlu 401
 DB 919 CCAATCTCTACTTGTAAACGAAATAATGTTGGATGTTGATGCAGATGCCAATGTACCGAA 978
 QY 402 GluAspSerGlySerArgLysLysIleThrCysGluCysThrLysProAspSerTyr 421
 DB 979 GAAGATTCAGGTAGCAACGAAAGAAATCACAATGTTGAATGTTACTAAACCTGATTTCT 1038
 QY 422 ProLeuPheAspGlyIlePheCysSer 430
 DB 1039 CCATCTTTTCGATGTTATTTTCTGCAGT 1065
 RESULT 11
 AAX25587
 ID AAX25587 standard; cDNA; 1088 BP.
 XX
 AC AAX25587;
 XX
 DT 02-AUG-1999 (first entry)
 XX
 DE Merozoite surface protein MSP-1-42 cDNA.
 XX
 KW MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 KW transgenic animal; ss.
 XX
 OS Plasmodium falciparum.
 XX
 FH Key Location/Qualifiers
 CDS 1..1086
 FT /*tag= a
 XX
 PN W09920766-A2.
 XX
 PD 29-APR-1999.
 XX
 PF 20-OCT-1998; 98WO-US22225.
 XX
 PR 15-MAY-1998; 98US-0085649.
 PR 20-OCT-1997; 97US-0062592.

XX (GENZ) GENZYME TRANSGENICS CORP.
 XX Chen LH, Meade H;
 XX WPI: 1999-302742/25.
 DR P-PSDB; AAY05833.
 DR
 PT New modified recombinant nucleic acid sequences useful for producing
 PT malarial DNA vaccine
 PS Disclosure; Fig 2; 43pp; English.
 XX
 CC This nucleic acid encodes a 42 kDa C-terminal portion (see AAY05833)
 CC of malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
 CC important target for the development of a vaccine against
 CC Plasmodium falciparum. The 3' end of the sequence has been
 CC modified to include a 6xHis tag. The nucleic acid sequence has
 CC been modified (see AAX25586) according to a method of the invention
 CC in order to improve expression in mammalian host cells and in
 CC transgenic animals. In the modified coding sequence, 306 nucleotide
 CC positions have been replaced to lower the AT content (from 76 to
 CC 49.7%) and to eliminate 10 mRNA instability motifs. The encoded
 CC amino acid sequence is unaltered. In another modified sequence
 CC (see AAX25593), a signal peptide sequence has been added and two
 CC N-glycosylation sites eliminated. The invention provides modified
 CC recombinant nucleic acid sequences and methods for increasing the
 CC mRNA levels and protein expression of proteins that are difficult
 CC to express in cell culture systems, mammalian cell culture systems
 CC or in transgenic animals. The preferred difficult protein
 CC candidates for expression are those derived from lower organisms
 CC such as parasites, bacteria and viruses that have DNA coding
 CC sequences of high AT content or which have mRNA instability motifs
 CC or rare codons relative to the recombinant expression system to be
 CC used. The invention allows expression of MSP-1 protein in the milk
 CC of transgenic animals, and also provides a DNA vaccine comprising a
 CC vector containing the altered MSP-1-42 sequence.
 XX
 SO Sequence 1088 BP; 454 A; 139 C; 150 G; 345 T; 0 other;

Alignment Scores:
 Pred. No.: 4.83e-88 Length: 1088
 Score: 1161.50 Matches: 218
 Percent Similarity: 73.71% Conservative: 54
 Best Local Similarity: 59.08% Mismatches: 76
 Query Match: 51.05% Indels: 21
 DB: 20 Gaps: 2

US-10-057-531A-2 (1-431) x AAX25587 (1-1088)

QY 63 MetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeu 82
 Db 19 ATTGATAACATACATCTTCTAAATTTGAAATTAAGTAAATACAGGTTTATATATTTAAACACCTTTA 78
 QY 83 AlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIlePheThrPheAsnLeu 102
 Db 79 GCAGGTGTTATACAGTGTAAAAAACAATAGAAATACAGTTATACAGTTATTAATGTT 138
 QY 103 AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeu 122
 Db 139 AATGTTAAGGATATTTTAAATTCAGATTTTAAATACAGTGAATTTTCAAAATGTTTAA 198
 QY 123 GluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSer 142
 Db 199 GAATCAGATTTTAAATTCATTAAGATTTTAAATCAATCAAGTAAATATGTTGCAAGATCCA 258
 QY 143 PheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysIleLys 162
 Db 259 TATAAATTTCTTATAAGAAAAAGAGATAAATCTTAACGACGTTATATATATATTAAG 318
 QY 163 GluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLysVal 182
 Db 319 GATTCAATAGATACGGATATATAAATTTTGCAAAATGATGTTCTTGGATATATATAAATATTA 378

QY 183 LeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluLysGlu 202
 Db 379 TCCGAAAAATATAATCAGATTAGATTCAATTAATAAATATATC----- 423
 QY 203 LysPheProSerSerProProThrThrProProSerProAlaLysThrAspGluGlnLys 222
 Db 424 -----AACGACAAACAAGGT 438
 QY 223 LysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuVal 242
 Db 439 GAAATGAGAAATACCTTCCTTTTAAACAATATGTAGACCTTATATAAACACGTTAAT 498
 QY 243 AsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGlu 262
 Db 499 GATAAATGATTATTGTTGTAATTCATTAGAACCAAAAGTTCTAAATATATCATATGAG 558
 QY 263 LysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAspLys 282
 Db 559 AAATCAACCGTAGAAGTTAAATTAAGAACTTAATTAATAAACAATTTCAAGACAAA 618
 QY 283 IleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAspAsp 302
 Db 619 TTGCACATTTTAAAAAATAACAAATTCCTTGAATTCGTGATTTATCAACAGATTAT 678
 QY 303 ThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal---GlnAsnPhePro 321
 Db 679 AACCATATAACTATTGACAAAGTCTCTTAGTACAGGTATGGTTTGAATAATCTTGCT 738
 QY 322 AsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSerGln 341
 Db 739 AAAACCCGTTTATCTAATTTACTTTGGAACCTTGAAGGTATGTTAAACATTTCAAA 798
 QY 342 HisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGlu 361
 Db 799 CACCAATCGGTAAACAAACAAATGTCACAAATTTCTGGATGTTTCAGACATTTAGATGAA 858
 QY 362 ArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluLysAspLysCysValIleAsn 381
 Db 859 AGAGAAGAAATGTAATGTTTATTAATAATACAAACAAGGTGATAATGTTTGAATAAT 918
 QY 382 ProAsnProThrCysAsnGluAsnAsnGlyCysAspAlaAspAlaThrCysThrGlu 401
 Db 919 CCAATCCTACTCTGTAACGAAATAATGTTGGATGTTGATGCAGATGCCAAATGTACCGAA 978
 QY 402 GluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLysProAspSerTyr 421
 Db 979 GAAGATTTCAGTAGCAACGGAAGAAATCACATGTCATGTAATGTAACCTGATTCATTAT 1038
 QY 422 ProIlePheAspGlyIlePheCysSer 430
 Db 1039 CCACCTTCGATGTTATTTCTGCACT 1065
 RESULT 12
 ID AAV35363
 XX AAV35363 standard; DNA; 4940 BP.
 AC AAV35363;
 XX
 DT 23-SEP-1998 (first entry)
 XX
 DE P. falciparum gp190 DNA.
 XX
 KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
 KW monoclonal antibody; passive immunisation; parasite; ss.
 XX
 OS Plasmodium falciparum.
 XX
 FH Key Location/Qualifiers
 FT CDS 10..4929
 FT /*tag= a
 FT /product= gp190
 XX

XX PN WO9920774-A2.
 XX PD 29-APR-1999.
 XX PF 20-OCT-1998; 98WO-US22226.
 XX PR 15-MAY-1998; 98US-0085649.
 XX PR 20-OCT-1997; 97US-0062592.
 XX PA (GENZ) GENZYME TRANSGENICS CORP.
 XX PI Chen LH, Meade H;
 XX DR WPI; 1999-288313/24.
 XX DR P-PSDB; AAY09374.
 XX PT Modified malarial protein for use in anti-malarial vaccines
 XX PS Example; Fig 11; 35pp; English.
 XX CC This novel, modified nucleic acid encodes a 42 kDa C-terminal
 CC part (see AAY09374) of malaria merozoite surface protein Msp-1
 CC (Msp-1-42), an important target for the development of a vaccine
 CC against Plasmodium falciparum. The nucleic acid sequence has been
 CC modified compared to the native sequence of Msp-1-42 (see AAX56009)
 CC such that the AT content has been reduced and 10 mRNA instability
 CC motifs eliminated while maintaining the protein amino acid sequence.
 CC In addition, a sequence encoding a 15-amino acid beta-casein signal
 CC peptide has been added to the 5' end of the sequence, and N262Q
 CC and N181Q mutations have been introduced to eliminate
 CC N-glycosylation sites. These alterations allow Msp-1-42 to be
 CC expressed in the mammary gland (i.e. milk) of transgenic mice. The
 CC invention also provides a DNA vaccine comprising a vector containing
 CC an altered Msp-1-42 sequence.
 XX S0 Sequence 1140 BP; 353 A; 282 C; 290 G; 215 T; 0 other;

Alignment Scores:
 Pred. No.: 1.97e-87 Length: 1140
 Score: 1154.50 Matches: 217
 Percent Similarity: 73.17% Conservative: 53
 Best Local Similarity: 58.81% Mismatches: 78
 Query Match: 50.75% Indels: 21
 DB: 20 Gaps: 2

US-10-057-531A-2 (1-431) x AAX56021 (1-1140)

Qy 63 MetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeu 82
 Db 64 ATCGATACATCCCTGTCACAGATCGAGAACGAGTACGAGGTGCTGCTACCTGAAGCCCTG 123

Qy 83 AlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeu 102
 Db 124 GCAGGAGTCTACAGGAGCTGACAGAGCAGCTGGAGACACAGCTGACCTTCAACGTG 183

Qy 103 AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeu 122
 Db 184 AACCTGAAGGATATCTCTCAACAGCAGGTTCACACAGAGGAGGAACCTTCAAGAACCTGCTG 243

Qy 123 GluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSer 142
 Db 244 GAGACCGATCTGATCCCTACAGATCTGACACGACCACTACGTGGTCAAGATCCC 303

Qy 143 PheLysLeuLeuAsnSerGluLysAsnThrLeuLeuLysSerTyrLysTyrIleLys 162
 Db 304 TACAAGTCTCTGACACAGGAGAGAGATAGTTCCTGACGAGTTACATATCATCAAG 363

Qy 163 GluSerValGluAsnAspIleLysPheAlaGlnGluIleSerTyrTyrGluLysVal 182
 Db 364 GATAGCATTTGACCGGATATCAACTTCGCCCAACGATGCTCTGGGATACACTCAAGATCTG 423

Qy 183 LeuAlaLysTyrLysAspLeuGluSerIleLysLysValIleLysGluLysGlu 202

Db 424 TCCGAGAAGTACAGAGCGATCTGGATAGCATCAAGAAAGTACATC----- 468
 Qy 203 LysPheProSerSerProProThrThrProProSerProAlaLysThrAspGluGlnLys 222
 Db 469 -----AACGATAAGCAGGGA 483

Qy 223 LysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuVal 242
 Db 484 GAGAACGAGAAGTACCTGCCCTTCTGAAACACATCGAGACCTGTACAAGACCGCTCAAC 543

Qy 243 AsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGlu 262
 Db 544 GATAAGATTGATCTGTTCTGTCATCCACCTGAGGCGCAAGGTCTCGAGTACATATGAG 603

Qy 263 LysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAspLys 282
 Db 604 AAGAACACGTGGAGGTCAAGATCAAGAGCTGAATTTACCTGAAGACCATCCAGGATAG 663

Qy 283 IleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAspAsp 302
 Db 664 CTGGCCGATTTCAGAACAGCAACAACATTCCTCGGAATCGCGATCTCGAGCAGCATTAC 723

Qy 303 ThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal---GlnAsnPhePro 321
 Db 724 AACCAACAACCTGCTGACCAAGTTCCTGAGCACCGGAATGGTCTTCGAAACCTGGCC 783

Qy 322 AsnThrIleLeuSerLysLeuIleGluLysPheGlnAspMetLeuAsnIleSerGln 341
 Db 784 AAGACCGTCTGACCAACCTGCTGATGGAACCTGACGGGATGCTCGAGATCAGCCAG 843

Qy 342 HisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGlu 361
 Db 844 CACCAGTGTGTGAAGACAGTGTCCCGACAGACGCGATGCTTCAGACACCTGGATGAG 903

Qy 362 ArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluLysAspLysCysValGluAsn 381
 Db 904 AGGAGGAGTGCAGTGCCTGCTGAACCTACAGCAGGAGGAGATAAGTGTGTGGAAAC 963

Qy 382 ProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAlaThrCysThrGlu 401
 Db 964 CCCAATCCTACTTGTACGAGAACAACTGAGGATGCGATGCGATGCCAAGTGTACCCAG 1023

Qy 402 GluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLysProAspSerTyr 421
 Db 1024 GAGGATTCAGGAAGCACAGGAAGATCACCTGCGAGTGTACCAAGCCTGATTCTTAT 1083

Qy 422 ProLeuPheAspGlyIlePheCysSer 430
 Db 1084 CCATGTTTCGATGTTATTTCTGCACT 1110

RESULT 14
 AAX25593
 ID AAX25593 standard; cDNA; 1140 BP.
 XX AAX25593;
 XX 02-AUG-1999 (first entry)
 XX Merozoite surface protein MSP-1-42 modified cDNA.
 KW MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 KW transgenic animal; mutant; ss.
 XX Plasmodium falciparum.
 OS Synthetic.
 XX Key Location/Qualifiers
 FT CDS 1..1131
 FT sig_peptide 1..45 a
 FT /*tag= b

XX PD 26-OCT-2000.
 XX PF 20-APR-2000; 2000WO-CB01558.
 XX PR 20-APR-1999; 98GB-0009072.
 PR 13-MAY-1999; 99US-0311817.
 PR 25-MAY-1999; 99CA-2271451.
 XX PA (MEDI-) MEDICAL RES COUNCIL.
 XX PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaipibull C;
 XX WPI; 2001-015762/02.
 DR P-PSDB; AAB37610.
 XX PT Novel variants of the C-terminal fragment of Plasmodium merozoite
 PT surface protein-1, useful as vaccines for treating or preventing
 PT malaria -
 XX
 PS Example 5; Fig 15; 126pp; English.
 CC The present invention relates to non-natural variants of a C-terminal
 CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
 CC non-natural variants have reduced affinity for at least 1 antibody
 CC capable of blocking a second antibody that inhibits the proteolytic
 CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
 CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
 CC MSP-1.4.2, compared to natural MSP-1.1.9. The non-natural variants of the
 CC present invention are useful for immunising a mammal against malaria, and
 CC can be used to treat malaria. The present sequence is the MSP-133
 CC coding sequence.
 XX
 SQ Sequence 786 BP; 245 A; 164 C; 159 G; 218 T; 0 other;

Alignment Scores:

Pred. No.: 1.21e-44 Length: 786
 Score: 640.50 Matches: 129
 Percent Similarity: 65.22% Conservative: 51
 Best Local Similarity: 46.74% Mismatches: 75
 Query Match: 28.15% Indels: 21
 DB: 22 Gaps: 2

US-10-057-531A-2 (1-431) x AAC68978 (1-786)

QY 63 MetAspAsnIleuSerGlyPheGluAsnGluTyArgValIleTyLeuLysProLeu 82
 DB 19 ATCGATAACATTCTGTCTPAAGATTGAGAACGATACGAGGTCTTGTACTTGAAGCTCTG 78
 QY 83 AlaGlyValTyArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeu 102
 DB 79 GCGGTGTCTACAGATCCCTGAAGAGCACTGGAAACAACGTCATGACTTCAACGTT 138
 QY 103 AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyPheLeuAspValLeu 122
 DB 139 ACGCTGAAGCATTTTGAACCTCCAGATTCACAAAGAGAGAAACCTCAAGACGTTCTG 198
 QY 123 GluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyIleIleGluAspSer 142
 DB 199 GAGCTGACTTGTATCCATCAAGGATTGACTTCTTCAACTACGTTGTTAAGGCCCA 258
 QY 143 PheLysLeuLeuAsnSerGlnLysAsnThrLeuLysSerTyTyTyIleLys 162
 DB 259 TACAAGTCTCTGAACAGGAGAGAGACAAAGTTCTTCTCTCTTACAACTACATTAAAG 318
 QY 163 GluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyTyTyIleLysVal 182
 DB 319 GACTCCCTGACTATGATATCACTTTCGTAACGAGCTCTGGGTACTTACAAAGATCTTG 378
 QY 183 LeuAlaLysTyTyTyAspAspLeuGluSerIleLysValIleLysGluLysGlu 202
 DB 379 TCTGAGAACTACAGTCTGACTTGGATTCCATCAAGAACTATCATC----- 423

QY 203 LysPheProSerSerProThrThrProProSerProAlaLysThrAspGluGlnLys 222
 DB 424 -----AATGATAGCAAGGA 438
 QY 223 LysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrrAsnLeuVal 242
 DB 439 GAGAAATGAAAGTACCTGCTCATTCATTTGGAAGCAAGGCTTGCATACACTTGTAAAC 498
 QY 243 AsnLysIleAspAspTyrrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGlu 262
 DB 499 GATAGATCGATCTGTGCTCATTCATTTGGAAGCAAGGCTTGCATACACTTGTAAAC 558
 QY 263 LysAspGluAlaHisValLysIleThrLysLeuSerAspLysAlaIleAspAspLys 282
 DB 559 AAGTCTAACCTCGAGTCAAGATCAAGGAATTGAACCTACCTCAAGACCATTCAGGATAAG 618
 QY 283 IleAspLeuPheLysAsnProTyrrAspPheGluAlaIleLysLysLeuIleAspAsp 302
 DB 619 CTGGCTGATTTCAGGAAGAACCAACACTTCGTTGGTATCGCTGATTGTCCACTGATTAC 678
 QY 303 ThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal--GlnAsnPhePro 321
 DB 679 AACCAACAACACTGTTGACTAAGTTCCTGCTACCGGTATGGTTTCGAGAACTTGGCT 738
 QY 322 AsnThrIleLysSerLysLeuIleGlyLysPheGlnAspMetLeu 337
 DB 739 AAGACTGTCTCTCCAAACCTGTTGGATGCTAACTTCAGGGTATGTTG 786
 RESULT 17
 AAT80403
 ID AAT80403 standard; DNA; 354 BP.
 XX
 AC AAT80403;
 XX
 DT 25-MAR-1998 (first entry)
 XX
 DE PfMSPL(p19)A coding sequence.
 XX
 KW Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
 KW Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.
 XX
 OS Plasmodium falciparum.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..354
 FT /tag= a
 FT /product= "PfMSPL(p19)A with GPI anchoring sequence"
 FT /note= "sequence contains 2 in frame stop codons at
 FT the 3'-end"
 FT misc_feature 1..285
 FT /tag= b
 FT /note= "derived from P. falciparum MSP1 p19 sequence"
 FT misc_feature 286..354
 FT /tag= c
 FT /note= "glycosylphosphatidylinositol anchoring sequence"
 XX
 PN WO9730159-A2.
 XX
 PD 21-AUG-1997.
 XX
 PF 14-FEB-1997; 97WO-FR00291.
 XX
 PR 14-FEB-1996; 96FR-0001821.
 XX
 PA (INSP) INST PASTEUR.
 PA (UUNY) UNIV NEW YORK STATE.
 XX
 PI Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
 XX Roth C;
 XX WPI; 1997-425034/39.
 DR P-PSDB; AAW22592.

XX AAT80404;
 AC 25-MAR-1998 (first entry)
 DT PfMSPl(p19)S coding sequence.
 DE Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
 XX Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.
 KW Chimeric - Plasmodium vivax.
 OS Chimeric - Plasmodium falciparum.
 XX Key Location/Qualifiers
 FT 1..387
 FT CDS /tag= a
 FT /product= PfMSPl(p19)S
 FT /note= "sequence contains 2 in frame stop codons at
 FT the 3' end"
 FT sig_peptide 1..57
 FT mat_peptide 58..381
 FT /tag= c
 FT misc_feature 1..96
 FT /tag= d
 FT /note= "sequence derived from P. vivax MSP1"
 FT misc_feature 97..102
 FT /tag= e
 FT /note= "sequence derived from generated restriction
 FT enzyme site"
 FT misc_feature 103..381
 FT /tag= e
 FT /note= "sequence derived from P. falciparum p19 coding
 FT sequence"
 XX WO9730159-A2.
 XX 21-AUG-1997.
 XX 14-FEB-1997; 97WO-FR00291.
 XX 14-FEB-1996; 96FR-0001821.
 XX (INSP) INST PASTEUR.
 XX (UUNY) UNIV NEW YORK STATE.
 XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
 XX Roth C;
 XX WPI; 1997-425034/39.
 XX P-PSDB; AAW22592.
 XX Recombinant protein containing Plasmodium merozoite surface
 PT protein-1 p42 fragment - useful in antimalarial vaccines, also new
 XX antibodies for diagnosis and protein purification
 XX Disclosure; Fig 1C; 85pp; French.
 XX This is the nucleotide sequence encoding a chimeric protein comprising
 CC nucleotides encoding amino acids 1-32 of the Plasmodium vivax merozoite
 CC surface protein 1 (MSP1) linked to the nucleotide sequence encoding the
 CC 19 kD C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
 CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
 CC The recombinant protein can be used for the production of anti-malarial
 CC vaccines, where the p19 fragment provides a high level of protective
 CC immunity since it includes epitopes not presented in the p42 fragment.
 XX Sequence 387 BP; 116 A; 94 C; 96 G; 81 T; 0 other;
 SQ

Alignment Scores:

Prod. No.: 1.59e-35 Length: 387
 Score: 527.00 Matches: 91
 Percent Similarity: 95.83% Conservative: 1

Best Local Similarity: 94.79% Mismatches: 4
 Query Match: 23.16% Indels: 0
 DB: 18 Gaps: 0
 US-10-057-531A-2 (1-431) x AAT80404 (1-387)
 QY 335 AspMetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGly 354
 DB 94 GAGCAATTCACATCTCGCAGCAGCCCAATCGTGAAAAACAATGTCCCGAGAACTCTGGC 153
 QY 355 CysPheArgHisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGlu 374
 DB 154 TGTTCAGACACTTGGCAGCAGAGAGAGAGTGTAAATGTCTGCTGAACACAGGAG 213
 QY 375 GlyAspLysCysValGluAsnProAsnProThrCysAsnGluAsnGlnGlyCysAsp 394
 DB 214 GCGCAACAGTGGTGAGAACCCCAACCCGACCTGTACGAGAACCAACCGCGCTGTGAC 273
 QY 395 AlaAspAlaThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGlu 414
 DB 274 GCAGACGCCCAATGCACCGCAGGAGGACTCGGCAGCAGCAAGAAATCAGCTGTGAG 333
 QY 415 CysThrLysProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
 DB 334 TGTACCAACCGGACTCGTACCGCTGTTCGACGGCATCTTCTGCAGC 381
 RESULT 20
 AAT94549
 ID AAT94549 standard; DNA; 387 BP.
 XX AC AAT94549;
 XX DT 25-MAR-1998 (first entry)
 XX DE PfMSPl(p19)S coding sequence.
 KW Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
 KW Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.
 XX Chimeric - Plasmodium vivax.
 OS Chimeric - Plasmodium falciparum.
 XX Key Location/Qualifiers
 FT 1..387
 FT CDS /tag= a
 FT /product= PfMSPl(p19)S
 FT /note= "sequence contains 2 in frame stop codons at
 FT the 3' end"
 FT sig_peptide 1..57
 FT /tag= b
 FT mat_peptide 58..381
 FT /tag= c
 FT misc_feature 1..96
 FT /tag= d
 FT /note= "sequence derived from P. vivax MSP1"
 FT misc_feature 97..102
 FT /tag= e
 FT /note= "sequence derived from generated restriction
 FT enzyme site"
 FT misc_feature 103..381
 FT /tag= e
 FT /note= "sequence derived from P. falciparum p19 coding
 FT sequence"
 XX WO9730159-A2.
 XX 21-AUG-1997.
 XX 14-FEB-1997; 97WO-FR00290.
 XX 14-FEB-1996; 96FR-0001822.
 XX (INSP) INST PASTEUR.

PA (UUNY) UNIV NEW YORK STATE.
 XX Barnwell, JW, Longacre-Andre S, Mendis K, Nato F;
 PI Roth C;
 XX WPI: 1997-425033/39.
 DR P-PSDB: AAW36102.
 XX
 XX Recombinant protein containing the merozoite surface protein-1 p19
 PT fragment - useful in anti-malarial vaccines, diagnosis and protein
 PT purification
 XX
 XX Disclosure; Fig 1C; 85pp; French.
 XX
 CC This is the nucleotide sequence encoding a chimeric protein comprising
 CC nucleotides encoding amino acids 1-32 of the Plasmodium vivax merozoite
 CC surface protein 1 (MSP1) linked to the nucleotide sequence encoding the
 CC 19 kD C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
 CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
 CC The recombinant protein can be used for the production of anti-malarial
 CC vaccines, where the p19 fragment provides a high level of protective
 CC immunity since it includes epitopes not presented in the p42 fragment.
 XX
 SQ Sequence 387 BP; 116 A; 94 C; 96 G; 81 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.59e-35 Length: 387
 Score: 527.00 Matches: 91
 Percent Similarity: 95.83% Conservatives: 1
 Best Local Similarity: 94.79% Mismatches: 4
 Query Match: 23.16% Indels: 0
 DB: 18 Gaps: 0

US-10-057-531A-2 (1-431) x AAT94549 (1-387)
 Qy 335 AspMetLeuAsnIleSerGlnHisGlnCysValLysGlnCysProGluAsnSerGly 354
 Db 94 GACGAATTCACATCTCCACACCAATGCGTGAATAAACAATGTCGGAAGACTCTGGC 153
 Qy 355 CysPheArgHisLeuAspGluArgGluGlnCysLysCysLeuLeuAsnTyrLysGlnGlu 374
 Db 154 TGTTCAGACACTGGACGAGAGAGAGAGTGAATGCTGCTGAACACAAACAGGAG 213
 Qy 375 GlyAspLysCysValGluAsnProAsnProThrCysAsnGlnAsnAsnGlyLysCysAsp 394
 Db 214 GCGCAAGTGGTGGAGAACCCCAACCCGACCTGTAAACAGAGAACACGGCGCTGTGAC 273
 Qy 395 AlaAspAlaThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGlu 414
 Db 274 GCAGACGCCAAATGCACCGAGAGAGACTCGGCGACGACGACGCAAGAAATCACGTGTGAG 333
 Qy 415 CysThrLysProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
 Db 334 TGTACCAAAACCCGACTCGTACCCGCTGTTCACGCGCATCTTCTGCAGC 381

RESULT 21
 AAC68977
 ID AAC68977 standard; DNA; 333 BP.
 XX
 AC AAC68977;
 XX
 DT 27-FEB-2001 (first entry)
 XX
 DE Merozoite surface protein-119 coding sequence.
 XX
 KW Merozoite surface protein; protazoacide; vaccine; malaria; ss.
 OS Plasmodium falciparum.
 XX
 PN WO200063245-A2.
 XX
 PD 26-OCT-2000.
 XX

PF 20-APR-2000; 2000WO-GB01558.
 XX
 PR 20-APR-1999; 99GB-0009072.
 PR 13-MAY-1999; 99US-0311817.
 PR 25-MAY-1999; 99CA-2271451.
 XX
 PA (MEDI-) MEDICAL RES COUNCIL.
 XX
 PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaiipibull C;
 PI WPI: 2001-015762/02.
 DR P-PSDB: AAB37609.
 XX
 PT Novel variants of the C-terminal fragment of Plasmodium merozoite
 PT surface protein-1, useful as vaccines for treating or preventing
 PT malaria
 XX
 PS Example 5; Fig 15; 126pp; English.
 XX
 CC The present invention relates to non-natural variants of a C-terminal
 CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
 CC non-natural variants have reduced affinity for at least 1 antibody
 CC capable of blocking a second antibody that inhibits the proteolytic
 CC cleavage of Plasmodium MSP-1_4_2, and has the same affinity for at least
 CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
 CC MSP-1_4_2, compared to natural MSP-1_1_9. The non-natural variants of the
 CC present invention are useful for immunising a mammal against malaria, and
 CC can be used to treat malaria. The present sequence is the MSP-119
 CC coding sequence.
 XX
 SQ Sequence 333 BP; 107 A; 75 C; 73 G; 78 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.95e-35 Length: 333
 Score: 525.00 Matches: 92
 Percent Similarity: 93.20% Conservatives: 4
 Best Local Similarity: 89.32% Mismatches: 3
 Query Match: 23.08% Indels: 4
 DB: 22 Gaps: 1

US-10-057-531A-2 (1-431) x AAC68977 (1-333)
 Qy 329 IleGluGlyLysPheGlnAspMetLeuAsnIleSerGlnHisGlnCysValLysGln 348
 Db 22 ATTGAAGGTAGACAC-----AACATTGCCCAACACCAATGCGTTAAGAACAA 69
 Qy 349 CysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGlnCysLysCysLeu 368
 Db 70 TGTCCACAAACTCCGCGATGTTTCAGACATCTGCACGAGAGAGAATGTAAGTGTCTG 129
 Qy 369 LeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsnProThrCysAsnGlu 388
 Db 130 TTGAACCTACAGCAGAGAGGTGATAAGTGTGTGAGAACCCAAACCTACCTGTTAACGAG 189
 Qy 389 AsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGluGluAspSerGlySerArg 408
 Db 190 AACACCGTGGATCGGCGCTGACGCTAAGTGCACCAAGAGACTCTGTTCTTAACGGA 249
 Qy 409 LysLysIleThrCysGluCysThrLysProAspSerTyrProLeuPheAspGlyIlePhe 428
 Db 250 AAGAAGATTACTTGGGAATGTAAGTGTGTGAGAACCCAAACCTACCTGTTTCGATGTAATCTC 309
 Qy 429 CysSerSer 431
 Db 310 TGTTCCTCC 318

RESULT 22
 AAD22459
 ID AAD22459 standard; DNA; 309 BP.
 XX
 AC AAD22459;
 XX
 DT 12-FEB-2002 (first entry)


```
XX Plasmodium falciparum PfMSP1.19 insert in pUC105-01 vector.
DE PUC105-01 vector; immunostimulant; vaccine; immunisation; therapeutic;
KW C3d; immune response; ds.
XX Plasmodium falciparum.
OS WO200177324-A1.
XX 18-OCT-2001.
PN 09-APR-2001; 2001WO-GB01599.
PD 08-APR-2000; 2000GB-0008582.
XX (ADPR-) ADPROTECH LTD.
XX Steward M, Cox VF;
PI WPI; 2002-010909/01.
DR Novel variant DNA sequence useful in DNA vaccine, encodes naturally
XX occurring protein and comprises a sequence non-identical to naturally
XX occurring DNA sequence encoding the protein
XX Example 13; Page 65; 87pp; English.
PS The invention relates to a variant DNA sequence useful in DNA vaccines.
CC The DNA sequence encodes a naturally occurring protein such as C3d which,
CC by virtue of third base redundancy and other variations permissible
CC within an amino acid codon, is non-identical to the naturally occurring
CC DNA sequence encoding that protein. The DNA sequence is useful in a DNA
CC immunisation vector to encode one or more naturally occurring human or
CC non-human proteins with immunomodulatory properties. The DNA sequence
CC is useful for inducing an immune response to an antigen in a human or
CC animal. A pharmaceutical composition is useful for introducing a DNA
CC sequence encoding a naturally occurring protein into a human or animal
CC by administering the pharmaceutical composition into the human or animal,
CC where the administration results in a therapeutic effect on the human or
CC animal. The present sequence is Plasmodium falciparum PfMSP1.19 insert
CC in pUC105-01 vector
XX
SQ Sequence 309 BP; 94 A; 71 C; 69 G; 75 T; 0 other;

Alignment Scores:
Pred. No.: 309 Length: 309
Score: 521.00 Matches: 89
Percent Similarity: 97.8% Conservative: 3
Best Local Similarity: 94.68% Mismatches: 2
Query Match: 22.90% Indels: 0
DB: 24 Gaps: 0

US-10-057-531A-2 (1-431) x AAD22459 (1-309)
QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 7 AACATTGCCCAACACCAATCGCTTAAAGAGCAATGCCACAAACATCCCGATGTTTCAGA 66
QY 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
Db 67 CATCTGGACGAGAGAGAAGAATGTAAGTGTCTGTTGAACACTACAGCAGGAAGCTGATAAG 126
QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyCysAspAlaAspAla 397
Db 127 TGTGTTGAGACCAACCAACCTACCTGTACGAGAGAACACGGTGGACCGCTGACGCT 186
QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
Db 187 AAGTGCACCGAAGAGACTCTGTTCTTAACGGAAGAAAGATTACTTGCAGATCTACTAAG 246
QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
|||||
```

```
Db 247 CCAGACTCTTACCCTTTGTTGATGGAATCTTCTGTCTTCC 288
RESULT 23
AAD22460
ID AAD22460 standard; DNA; 3147 BP.
XX
XX AC AAD22460;
XX
DT 12-FEB-2002 (first entry)
XX
DE PfMSP1.19-human C3d3 DNA coding sequence from pVK104-01 vector.
XX
KW Human; immunostimulant; vaccine; immunisation; therapeutic; C3d;
XX immune response; pVK104-01 vector; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Plasmodium falciparum.
OS Chimeric - Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 1..3147
FT /tag= a
FT /product= "Human C3d3-PfMSP1.19 protein"
FT /note= "CDS does not include stop codon"
FT /partial
FT sig_peptide 1..72
FT /tag= b
FT mat_peptide 73..3147
FT /tag= c
FT /product= "Mature human C3d3 protein"
FT /note= "Antigen sequence"
FT misc_feature 73..369
FT /tag= d
XX WO200177324-A1.
XX 18-OCT-2001.
XX
XX 09-APR-2001; 2001WO-GB01599.
XX
XX 08-APR-2000; 2000GB-0008582.
XX
XX (ADPR-) ADPROTECH LTD.
XX
XX Steward M, Cox VF;
XX
XX WPI; 2002-010909/01.
XX
XX Novel variant DNA sequence useful in DNA vaccine, encodes naturally
XX occurring protein and comprises a sequence non-identical to naturally
XX occurring DNA sequence encoding the protein
XX Claim 16; Page 66-68; 87pp; English.
XX
XX The invention relates to a variant DNA sequence useful in DNA vaccines.
XX The DNA sequence encodes a naturally occurring protein such as C3d which,
XX by virtue of third base redundancy and other variations permissible
XX within an amino acid codon, is non-identical to the naturally occurring
XX DNA sequence encoding that protein. The DNA sequence is useful in a DNA
XX immunisation vector to encode one or more naturally occurring human or
XX non-human proteins with immunomodulatory properties. The DNA sequence
XX is useful for inducing an immune response to an antigen in a human or
XX animal. A pharmaceutical composition is useful for introducing a DNA
XX sequence encoding a naturally occurring protein into a human or animal
XX by administering the pharmaceutical composition into the human or
XX animal. The present sequence is Plasmodium falciparum PfMSP1.19-human
XX C3d3 DNA coding sequence from vector pVK104-01.
XX
XX Sequence 3147 BP; 869 A; 754 C; 834 G; 690 T; 0 other;

Alignment Scores:
Pred. No.: 6.41e-34 Length: 3147
```

Score: 521.00 Matches: 89
Percent Similarity: 97.87% Conservative: 3
Best Local Similarity: 94.68% Mismatches: 2
Query Match: 22.90% Indels: 0
DB: Gaps: 0

US-10-057-531A-2 (1-431) x AAD22460 (1-3147)

Qy 338 AsnIleSerGlnHisGlnCysValIysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 73 AACATGCCCAACACCACCGTTAAGAGCAATGTCCACAAACTCCGGATGTTTCAGA 132

Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAsPlys 377
Db 133 CATCTGGACGAGAGAGAGAAATGTAAGTGTCTGTGAACACAAAGAGGAGGTAAG 192

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
Db 193 TGTCTTGAGAACCCCAACCTCTGTAAACGAGAACCGTGGATCGGACGCTGACGCT 252

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
Db 253 AAGTGCACCGAAGAAAGACTCTGGTTCTTAACGGAAGAAGATTACTTCGGAATGTACTAAG 312

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
Db 313 CCAGACTCTTACCCTTTGTTTCGATGGAATCTCTGTTCTTCC 354

RESULT 24
AAD22461
ID AAD22461 standard; DNA: 3147 BP.
XX
AC AAD22461;
XX
DT 12-FEB-2002 (first entry)
XX
DE Human C3d3 DNA-PfMSP1.19 coding sequence from pVK104-02 vector.
XX
KW Human; Immunostimulant; vaccine; immunisation; therapeutic; C3d;
KW Immune response; pVK104-02 vector; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Plasmodium falciparum.
OS Chimeric - Unidentified.

Key Location/Qualifiers
CDS 1..3147
FT sig_peptide 1..72
FT mat_peptide 73..3147
FT /product= "Mature human C3d3-PfMSP1.19 protein"
FT /tag= d
FT /note= "Antigen sequence"

WO200177324-A1.
XX
XX 18-OCT-2001.
XX
XX 09-APR-2001; 2001WO-GB01599.
XX
XX 08-APR-2000; 2000GB-0008582.
XX
XX (ADPR-) ADPROTECH LTD.
XX
XX Steward M, Cox VF;
XX
XX WPI; 2002-010909/01.

XX Novel variant DNA sequence useful in DNA vaccine, encodes naturally
PT occurring protein and comprises a sequence non-identical to naturally
PT occurring DNA sequence encoding the protein
XX
PS Claim 16; Page 69-71; 87pp; English.
XX
CC The invention relates to a variant DNA sequence useful in DNA vaccines.
CC The DNA sequence encodes a naturally occurring protein such as C3d which,
CC by virtue of third base redundancy and other variations permissible
CC within an amino acid codon, is non-identical to the naturally occurring
CC DNA sequence encoding that protein. The DNA sequence is useful in a DNA
CC immunisation vector to encode one or more naturally occurring human or
CC non-human proteins with immunomodulatory properties. The DNA sequence
CC is useful for inducing an immune response to an antigen in a human or
CC animal. A pharmaceutical composition is useful for introducing a DNA
CC sequence encoding a naturally occurring protein into a human or animal
CC by administering the pharmaceutical composition into the human or animal,
CC where the administration results in a therapeutic effect on the human or
CC animal. The present sequence is Plasmodium falciparum PfMSP1.19-human
CC C3d3 DNA coding sequence from vector pVK104-02.
XX
SQ Sequence 3147 BP; 869 A; 754 C; 834 G; 690 T; 0 other;

Alignment Scores:
Pred. No.: 6,41e-34 Length: 3147
Score: 521.00 Matches: 89
Percent Similarity: 97.87% Conservative: 3
Best Local Similarity: 94.68% Mismatches: 2
Query Match: 22.90% Indels: 0
DB: Gaps: 0

US-10-057-531A-2 (1-431) x AAD22461 (1-3147)

Qy 338 AsnIleSerGlnHisGlnCysValIysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 2845 AACATGCCCAACACCACCGTTAAGAGCAATGTCCACAAACTCCGGATGTTTCAGA 2904

Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAsPlys 377
Db 2905 CATCTGGACGAGAGAGAAATGTAAGTGTCTGTGAACACAAAGAGGAGGTAAG 2964

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
Db 2965 TGTCTTGAGAACCCCAACCTCTGTAAACGAGAACCGTGGATCGGACGCTGACGCT 3024

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
Db 3025 AAGTGCACCGAAGAGACTCTGGTTCTTAACGGAAGAAGATTACTTCGGAATGTACTAAG 3084

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
Db 3085 CCAGACTCTTACCCTTTGTTTCGATGGAATCTCTGTTCTTCC 3126

RESULT 25
AAD22462
ID AAD22462 standard; DNA: 309 BP.
XX
AC AAD22462;
XX
DT 12-FEB-2002 (first entry)
XX
DE Plasmodium falciparum MSP1.19 (PfMSP1.19) mutant DNA.
XX
KW Immunostimulant; vaccine; immunisation; therapeutic; immune response;
KW MSP1.19 gene; mutant; ds.
XX
OS Plasmodium falciparum.
OS Synthetic.
XX
XX Key Location/Qualifiers
FT misc_feature 40..42
FT /tag= a

```
FT /note= "Mutation occurs at this positions which
FT results in an amino acid change from Cys to Ile"
FT 88..90
FT *tag= b
FT /note= "Mutation occurs at this positions which
FT results in an amino acid change from Cys to Trp"
XX
XX WO200177324-A1.
XX
XX 18-OCT-2001.
XX
XX 09-APR-2001; 2001WO-GB01599.
XX
XX 08-APR-2000; 2000GB-0008582.
XX
XX (ADPR-) ADPROTECH LTD.
XX
XX Steward M, Cox VF;
XX
XX WPI; 2002-010909/01.
XX
XX Novel variant DNA sequence useful in DNA vaccine, encodes naturally
XX occurring protein and comprises a sequence non-identical to naturally
XX occurring DNA sequence encoding the protein
XX
XX Example 13; Page 72; 87pp; English.
XX
XX The invention relates to a variant DNA sequence useful in DNA vaccines.
XX The DNA sequence encodes a naturally occurring protein such as C3d which,
XX by virtue of third base redundancy and other variations permissible
XX within an amino acid codon, is non-identical to the naturally occurring
XX DNA sequence encoding that protein. The DNA sequence is useful in a DNA
XX immunisation vector to encode one or more naturally occurring human or
XX non-human proteins with immunomodulatory properties. The DNA sequence
XX is useful for inducing an immune response to an antigen in a human or
XX animal. A pharmaceutical composition is useful for introducing a DNA
XX sequence encoding a naturally occurring protein into a human or animal,
XX by administering the pharmaceutical composition into the human or animal,
XX where the administration results in a therapeutic effect on the human or
XX animal. The present sequence is Plasmodium falciparum PfMSP1.19 mutant
XX DNA used in the exemplification of the invention.
XX
XX Sequence 309 BP; 95 A; 71 C; 69 G; 74 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 2.2e-33 Length: 309
XX Score: 500.00 Matches: 87
XX Percent Similarity: 95.74% Conservative: 3
XX Best Local Similarity: 92.55% Mismatches: 4
XX Query Match: 21.98% Indels: 0
XX DB: 24 Gaps: 0
XX
XX US-10-057-531A-2 (1-431) x AAD22462 (1-309)
XX
QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
DB 7 AACATTGCCACACCAATCGTTAAGAGCAAAATTCACAAACATCCGGATCTTCAGA 66
QY 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
DB 67 CATCTGGACGAGAGAGAAGAAATGGAATGCTCTTGAACCTACAGCAGGAAGCTGATAAG 126
QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyCysAspAlaAspAla 397
DB 127 TGTGTTGAGAACCCAAACCTACTGTAAAGCAGAACACCGTGGATGCGACGCTGACGCT 186
QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
DB 187 AAGTCGACCGAACAAGACTCTGTTCTTAAACGGAAGAAGATTACTTGCGAATGCTACTAAG 246
QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
DB 247 CCAGACTCTTACCCTTTGTTTCGATGGAATCTTCTGTCTTCC 288
```

```
RESULT 26
AAD22463
ID AAD22463 standard; DNA; 3147 BP.
XX
XX AAD22463;
XX
XX 12-FEB-2002. (first entry)
XX
XX Mutant PfMSP1.19-human C3d3 DNA coding sequence from pVK104-03 vector.
XX
XX Human; immunostimulant; vaccine; immunisation; therapeutic; C3d;
XX immune response; pVK104-03 vector; ds.
XX
XX Chimeric - Homo sapiens.
XX Chimeric - Plasmodium falciparum.
XX Chimeric - Unidentified.
XX
XX Key Location/Qualifiers
XX CDS 1..3147
XX FT /*tag= a
XX FT /product= "Human C3d3-PfMSP1.19 protein"
XX FT /note= "CDS does not include stop codon"
XX FT /partial
XX FT sig_peptide 1..72
XX FT /*tag= b
XX FT mat_peptide 73..3147
XX FT /*tag= c
XX FT /product= "Mature human C3d3 protein" ;
XX FT misc_feature 73..369
XX FT /*tag= d
XX FT /note= "Antigen sequence"
XX WO200177324-A1.
XX
XX 18-OCT-2001.
XX
XX 09-APR-2001; 2001WO-GB01599.
XX
XX 08-APR-2000; 2000GB-0008582.
XX
XX (ADPR-) ADPROTECH LTD.
XX
XX Steward M, Cox VF;
XX
XX WPI; 2002-010909/01.
XX
XX Novel variant DNA sequence useful in DNA vaccine, encodes naturally
XX occurring protein and comprises a sequence non-identical to naturally
XX occurring DNA sequence encoding the protein
XX
XX Claim 16; Page 73-75; 87pp; English.
XX
XX The invention relates to a variant DNA sequence useful in DNA vaccines.
XX The DNA sequence encodes a naturally occurring protein such as C3d which,
XX by virtue of third base redundancy and other variations permissible
XX within an amino acid codon, is non-identical to the naturally occurring
XX DNA sequence encoding that protein. The DNA sequence is useful in a DNA
XX immunisation vector to encode one or more naturally occurring human or
XX non-human proteins with immunomodulatory properties. The DNA sequence
XX is useful for inducing an immune response to an antigen in a human or
XX animal. A pharmaceutical composition is useful for introducing a DNA
XX sequence encoding a naturally occurring protein into a human or animal,
XX by administering the pharmaceutical composition into the human or animal,
XX where the administration results in a therapeutic effect on the human or
XX animal. The present sequence is Plasmodium falciparum mutant
XX PfMSP1.19-human C3d3 DNA coding sequence from vector pVK104-03.
XX
XX Sequence 3147 BP; 870 A; 754 C; 834 G; 689 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 3.65e-32 Length: 3147
XX Score: 500.00 Matches: 87
```


The invention relates to recombinant multivalent proteins (I) that stimulate an immune response to Plasmodium vivax. (I) comprises antigenic determinants, fragments or conservative substitutions, derived from more than one stage in a life cycle of a Plasmodium vivax parasite. (I) is useful as a vaccine for stimulating an immune response, specifically a protective immune response that confers increased resistance to infection by Plasmodium parasites, such as P. vivax. (I) is especially useful in the treatment, prevention and reduction of malarial infection, as research or diagnostic reagents for the detection of Plasmodium species in a biological sample, and for conferring immunity against multiple stages of the malarial parasite. The antibodies produced are useful for the detection or measurement of antigenic epitopes derived from one or more stages in a life cycle of a parasite, particularly P. vivax. The vaccine comprising the recombinant proteins, is cost-effective, health-promoting intervention for controlling, preventing or treating the incidence of malaria. The present sequence represents a synthetic gene vivac2 encoding the recombinant protein vivac2p, a multivalent and

CC multistage vaccine against P. vivax.

XX Sequence 6101 BP; 2077 A; 1168 C; 1534 G; 1322 T; 0 other;

Alignment Scores:

Pred. No.: 8.57e-14 Length: 6101
 Score: 284.50 Matches: 85
 Percent Similarity: 40.00% Conservative: 55
 Best Local Similarity: 24.29% Mismatches: 106
 Query Match: 12.51% Indels: 104
 DB: 22 Gaps: 10

US-10-057-531A-2 (1-431) x AAH47055 (1-6101)

QY 101 AsnLeuAsnLeuAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAsp 120
 DB 5295 AATTTGGACAGCAATGATCTACTAGTTAAAGAAAGATTAAAGGGCAATTTATATGGAT 5354
 QY 121 Val-----LeuGluSerAspLeuMetGlnPheLysHisIleSe 133
 DB 5355 TTGTAAATTAATTTGCGGTAAATATAGACCGCAGATATACAGATGGATTCGAGAATG 5414
 QY 133 rSerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnTh 153
 DB 5415 GGGAGGGGATTACGTGCTCAGAA-----TTGCCACAGAGTGCAGAAA----- 5457
 QY 153 rLeuLeuLysSerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPhe----- 171
 DB 5458 -----CTGAAGAGAAATGTGATGAGAAAATCAATTACTACTGA 5495
 QY 172 -----AlaGlnGluGlyIleSerTyrTyrGluLysValLe 183
 DB 5496 TAAAAAGTATGTAAGGTACCACCATGTCAAAATCGGTGTAATCATATGATCAATGGAT 5555
 QY 183 uAlaLysTyrLysAsp-----AspLeuGluSerIleLysLysValIleLysGluLuluy 201
 DB 5556 ACCGAGAAAATAAATCAATGGGTGCTGTCGAATAAATCAATGAAGTGAATAAACCGC 5615
 QY 201 sGluLysPheProSerProProThrProProSerProAlaLysThrAspGluG1 221
 DB 5616 AGAAAAGGTTACAGCGCGAGGTATCGTAACCTTATGATATATAAAGCAGGAGTTAGA 5675
 QY 221 nLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLe 241
 DB 5676 TGAATTTAAC-----GAGGTGGCTTTTGAGAATGA 5705
 QY 241 uValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnVa 261
 DB 5706 AATTACAAACGTGATGTCATATATTGAGTTATGCGTT-----TGTTCCTG 5753
 QY 261 lGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAspAs 281
 DB 5754 TGAAGAG-----GCTAAAAAATACTCAGGAAGTTGTGACAAA 5792
 QY 281 pLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAs 301
 DB 5793 TGTGGACAATGCTGCTAAATCTCCATGGACTATGAGC----- 5829
 QY 301 pasPthrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPhePr 321
 DB 5829 ----- 5829
 QY 321 oAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSerG1 341
 DB 5830 -----TCCGA 5834
 QY 341 nHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLysAspG1 361
 DB 5835 GCACACATGTATAGACACCAATGTGCTCATATGTCAGCCTGCTATAGTACTTGGACGG 5894
 QY 361 uArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAs 381
 DB 5895 AACGGAAGATGGAGTGTGTTGTTACCTTTAAAGAAAGGCGGCAAGTGTGTCGCAGC 5954

QY 381 nProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrG1 401
 DB 5955 ATCGAATGTGACTTGAAGATAACAATGGTGGTGGCCCTGAAGCTGAATGTAAAT 6014
 QY 401 uGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLysProAspSerTy 421
 DB 6015 GACGGACAGC-----AATAAATCGTCTCTAAATGTAATAAGAGGTCTCTGA 6062
 QY 421 rProLeuPheAspGlyIlePheCysSer 430
 DB 6063 GCCACCCCTTTGAGGGAGTTTCTGTAGC 6090

RESULT 31

AAH47056
 ID AAH47056 standard; DNA; 5643 BP.
 XX
 AC AAH47056;
 XX
 DT 29-OCT-2001 (first entry)
 XX
 DE Synthetic gene ViVac2bias nucleotide sequence.
 XX
 KW Multivalent protein; immune response; Plasmodium vivax; parasite;
 KW protozoacide; vaccine; malaria; recombinant; ViVacip; ds.
 XX
 OS Synthetic.
 OS Plasmodium vivax.
 PN WO20015181-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 29-JAN-2001; 2001WO-US02937.
 XX
 PR 31-JAN-2000; 2000US-0179213.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Lal AA, Xiao L, Zhou Z;
 DR WPI: 2001-514557/56.
 DR P-PSDB; AAB85698.

New recombinant multivalent protein comprising antigenic determinants derived from more than one stage in a life cycle of Plasmodium vivax, useful as a vaccine for treating, preventing and reducing malarial infection

Claim 9; Page 56-59; 59pp; English.

The invention relates to recombinant multivalent proteins (I) that stimulate an immune response to Plasmodium vivax. (I) comprises antigenic determinants, fragments or conservative substitutions, derived from more than one stage in a life cycle of a Plasmodium vivax parasite. (I) is useful as a vaccine for stimulating an immune response, specifically a protective immune response that confers increased resistance to infection by Plasmodium parasites, such as P. vivax. (I) is especially useful in the treatment, prevention and reduction of malarial infection, as research or diagnostic reagents for the detection of Plasmodium species in a biological sample, and for conferring immunity against multiple stages of the malarial parasite. The antibodies produced are useful for the detection or measurement of antigenic epitopes derived from one or more stages in a life cycle of a parasite, particularly P. vivax. The vaccine comprising the recombinant proteins, is cost-effective, health-promoting intervention for controlling, preventing or treating the incidence of malaria. The present sequence represents the nucleotide sequence of a synthetic gene ViVac2bias.

Sequence 5643 BP; 1584 A; 1197 C; 1503 G; 1359 T; 0 other;

Alignment Scores:

Pred. No.: 9.45e-14 Length: 5643

Db	5377	-----TCCGAGCACACTTTGTATCCGACACTAACGTGCCAGATAACGCTGCTTGTTAC	
Qy	357	ArgHisLeuAspLuarGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyasp	
Db	5428	AGATACTGGACCGAACTCAGGAGTCGGAGATGTTTGTGACTTTCAGGAGGAGGTGCT	
Qy	377	LysCysValcIuAsnProAsnProThrCysAsnGluAsnAsnGlyCysAspAlaAasp	
Db	5488	AAGTGTGTGCAGCTTCCAACCTGACTTGTGAAGGATAACAACGGTGGTGTGCTCCAGAG	
Qy	397	AlaThrCysThrcIucluaSpSerGlySerSerArgLysLysLleThrCysGluCysThr	
Db	5548	GCTGAGTGTAAAGATGACTGCTCC-----AACAGAATCGCTGTGAAGTGTA	
Qy	417	LysProAspSerTyrProLeuPheAspGlyllePheCysSer	
Db	5596	AAGGAGGGTCTCGAGCCACCTTCGAGGAGGTTTTCTGTCTCC	
RESULT 32			
ID	AAD22449	standard; DNA; 3252 BP.	
XX	AC	AAAD22449;	
DT	XX	12-FEB-2002 (first entry)	
DE	XX	PYMSP1.19-murine C3d chimeric gene coding sequence from pVK96-02 vector	
KW	MSPL	gene; immunostimulant; vaccine; immunisation; therapeutic; murine immune response; C3d; pVK96-02 vector; chimeric; ds.	
OS	Chimeric - Plasmodium yoelii.		
OS	Chimeric - Mus sp.		
OS	Chimeric - Unidentified.		
FH	Key	Location/Qualifiers	
FT	misc_feature	2842..3246	
FT	/tag=	a	
FT	/note=	"Antigen sequence"	
PN	WC200177324-A1.		
PD	18-OCT-2001.		
PF	09-APR-2001; 2001WO-CB01599.		
PR	08-APR-2000; 2000GB-0008582.		
PA	(ADPR-) ADPROTECH LTD.		
PI	Steward M, Cox VF;		
PI	WPI; 2002-010909/01.		
DR	Novel variant DNA sequence useful in DNA vaccine, encodes naturally occurring protein and comprises a sequence non-identical to naturally occurring DNA sequence encoding the protein		
PT	Claim 17; Page 55-57; 87pp; English.		
CC	The invention relates to a variant DNA sequence useful in DNA vaccines		
CC	The DNA sequence encodes a naturally occurring protein such as C3d which by virtue of third base redundancy and other variations permissible within an amino acid codon, is non-identical to the naturally occurring DNA sequence encoding that protein. The DNA sequence is useful in a DNA immunisation vector to encode one or more naturally occurring human or non-human proteins with immunomodulatory properties. The DNA sequence is useful for inducing an immune response to an antigen in a human or animal. A pharmaceutical composition is useful for introducing a DNA sequence encoding a naturally occurring protein into a human or animal by administering the pharmaceutical composition into the human or animal where the administration results in a therapeutic effect on the human		

CC animal. The present sequence is Plasmodium yoelii MSP1 (PymSP1.19)
 CC antigen sequence and murine C3d chimeric coding sequence from pVK96-02
 CC vector.

XX
 SQ Sequence 3252 BP; 864 A; 786 C; 858 G; 744 T; 0 other;

Alignment Scores:

Pred. No.: 5,88e-14 Length: 3252
 Score: 282.50 Matches: 119
 Percent Similarity: 39.76% Conservative: 83
 Best Local Similarity: 23.43% Mismatches: 175
 Query Match: 12.42% Indels: 131
 DB: 24 Gaps: 21

US-10-057-531A-2 (1-431) x AAD22449 (1-3252)

QY 22 AlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 41
 DB 1819 GCGCTCGCGCAGTACACAGCGAGCTGGCGAGATCAACAAGACCTCAATATGACGCTCACT 1878
 QY 42 AspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMetAlaIleSerVal 61
 DB 1879 TTTCAC---TTGCCATCATCCGGAGC-----GCGGAGGTGGGAGCGGGGGCGGT 1929
 QY 62 ThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysPro 81
 DB 1930 ACCTCGGATCTACCCCGCAGGC---TCTGGGAACACAGATGATTGGCATGACACCA 1986
 QY 82 ----LeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThr 99
 DB 1987 ACAGTCATTGGCTGGCTACACTTACCTGGACACCGACAGCTGGGAGAG----- 2034
 QY 100 PheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeu 119
 DB 2035 -----TTCCGATAGACAGAGGCAAGGCCCTCG 2064
 QY 120 AspValLeuGluSerAspLeuMetGln-----PheLysHisIleSerSerAsnGlu 136
 DB 2065 GAGCTCATCAAGAAAGGTATACCCAGCAGCTGGCGCTTCAACAGCCAGCTCT----- 2118
 QY 137 TyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLys 156
 DB 2119 -----GCCTATGCTGCCTTCAACAACCGCCGCCAGCACCTGGCTGACA 2163
 QY 157 SerTyrLys----- 159
 DB 2164 GCCTAGCTGGTCAAGGTCTTCTCTAGCTGCCAACCTCATCGCCATCGACTCTCACGTC 2223
 QY 160 -----TyrIleLysGluSerValGluAsnAspIleLysPheAla 172
 DB 2224 CTGTGTGGGCTGTTAAATGTTGATTCTGGAGAAACAGACGCGGATGGTGTCTTTTCAG 2283
 QY 173 GlnGluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSer 192
 DB 2284 GAGGATGGCCCGGTGATTCAACAAGAAATGATTGGTGGCTTCGGAACCCCAAGGAGGA 2343
 QY 193 -----IleLysLysValIleLysGluGluLysGluLysPhePro 205
 DB 2344 GATGTCTACTACAGCCTTCGTCTCTCATCGCATCGAGAACCCAGGCATCTGTGAG 2403
 QY 206 SerSerProProThrProProSerProAlaLysThrAspGlnLysLysGluSer 225
 DB 2404 GGGCAGGTCAATAGCTTCTCGGAGCATCAACAGGCGGGAGTATATTGAAGCCAGT 2463
 QY 226 LysPhe-----LeuProPheLeuThrAsnIleGluThrLeuTyrAsnLeuVal 242
 DB 2464 TACATGAACCTGCAGACCATACAGCTGGCCATTGCTGGGTATGCCCTGGCCCTCATG 2523
 QY 243 AsnLysIleAspAsp---TyrIleLeuAsnLeuLysAlaLysIleAsnAspCysAsn--- 260
 DB 2524 ACAAACTGGAGAACCTTACCTCGGCAAGTTCTGCAACACACACCAAGATCGGAACCGC 2583
 QY 261 ----ValGluLysAspGluAla-----HisValLysIleThrLysLeuSerAspLeuLys 277

DB 2584 TGGGAGGCGCTGACCGAGCAGCTCTACAACGTAGAGCCACATCTCTACGCCCTCTGGCC 2643
 QY 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
 DB 2644 CTGCTG-----CTGCTGAAA-----GACITTGACTCTGTGCCCCCT 2679
 QY 298 LeuIle----- 299
 DB 2680 GTAGTCGCTGGCTCAATGAGCANAAGATACTACGGAGCGGCTATGGCTCCACCCAGGCT 2739
 QY 300 -----AsnAspAsp 302
 DB 2740 ACCTTCATGGTATTTCAAGCCTTGGCCCAATATCAAAACAGATGTCCCTGACCATAGGAC 2799
 QY 303 ThrLysLysAspMet----- 307
 DB 2800 TTGAACATGGATGTGTCTCCACCTCCCGCAGCAGTGGATCTCACATTTGGCTCTATTGCT 2859
 QY 308 LeuGlyLysLeuLeuSerThrGlyLeuVal---GlnAsnPheProAsnThrIleIleSer 326
 DB 2860 TTGAACAACCTTGAACAAGTCTGGTTGGTAGGAGAGGTGAGTCTAAGAAGATTTTGGCT 2919
 QY 327 LysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSer---GlnHisGlnCysVal 345
 DB 2920 AAGATGCTGAACATGGACGCTATGGACTTGTGGTGTGACCCCTAAGCATGTTGTGTT 2979
 QY 346 ---LysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGlu 364
 DB 2980 GACACTAGAGACATCTTAAGAACGCTGGATGTTTCAGAGACGACACCGTACTGAAGAG 3039
 QY 365 CysLysCysLeuLeuAsnTyrLysGln---GluGlyAspLysCysValGluAsnProAsn 383
 DB 3040 TGGAGATGTTTGTGGGTTACAAGAAGGTGAGGGTAACACCTGCGTTGAGAACAACAAC 3099
 QY 384 ProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGluGluAsp 403
 DB 3100 CCTACTTGGGACATCAACAACGGTGGATGTGACCCCAACGCCCTCTTGTCAAAACGCTGA 3159
 QY 404 SerGlySerSerArgLysLysIleThrCysGluCysThrLysProAspSerTyrProLeu 423
 DB 3160 TCTACCGAAACCTCCAAAGAGATTTATTCACCTGTAAAGAACCAACCCCTAACGCCCTAC 3219
 QY 424 PheAspGlyIlePheCysSerSer 431
 DB 3220 TAGGAGGTGTTTCTGTTCTTCT 3243
 RESULT 33
 AAD21371
 ID AAD21371 standard; DNA; 870 BP.
 XX
 AC AAD21371;
 XX
 DT 28-JAN-2002 (first entry)
 XX
 DE Helicobacter pylori His-HOP38(-11) polypeptide encoding DNA.
 KW Polypeptide delivery system; immune stimulating complex; ISCOM;
 KW cholesterol; saponin; phospholipid; medicament; vaccine; therapy;
 KW Helicobacter pylori infection; HOP38; anti-bacterial; ds.
 XX
 OS Helicobacter pylori.
 FH Key Location/Qualifiers
 CDS 1..870
 FT /tag= a
 FT /product= "Helicobacter pylori His-HOP38(-11)
 FT polypeptide"
 PN WO200176625-A1.
 XX 18-OCT-2001.
 PD
 XX

PF 09-APR-2001; 2001WO-SE00799.

XX 12-APR-2000; 2000GB-0008879.

XX (ASTR) ASTRAZENECA AB.

XX Shapiro A, Sanyal G;

XX WPI; 2001-663016/76.

XX P-PSDB; AAEL3033.

XX Producing a polypeptide delivery system useful in a vaccine to treat

XX infection by mixing together the polypeptide, cholesterol, saponin, and

XX a phospholipid in presence of a nonionic detergent and a second

XX detergent

XX Disclosure; Page 38-39; 43pp; English.

XX The invention relates to producing a polypeptide delivery system

XX comprising an immune stimulating complex (ISCOM) coupled to a

XX polypeptide. The method involves mixing the polypeptide, cholesterol,

XX saponin and a phospholipid in the presence of a nonionic detergent

XX and a second detergent to form a solution and removing the detergent

XX from the mixture to form the ISCOM. The polypeptide delivery system

XX is used as a medicament for the manufacture of a vaccine for

XX administration to mammalian patients, to treat and prevent

XX Helicobacter pylori infection. The present sequence is a DNA encoding

XX His-HOP38(-11), a Helicobacter pylori polypeptide used in the

XX polypeptide delivery system.

XX Sequence 870 BP; 235 A; 170 C; 214 G; 251 T; 0 other;

SQ

Alignment Scores:

Pred. No.: 2,57e-14 Length: 870

Score: 278.50 Matches: 53

Percent Similarity: 96.36% Conservative: 0

Best Local Similarity: 96.36% Mismatches: 1

Query Match: 12.24% Indels: 1

DB: 22 Gaps: 1

US-10-057-531A-2 (1-431) x AAD21371 (1-870)

Qy 1 MethHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20

Db 1 ATGCACCATCATCATCATCATCTCTGGTCTGGTCCACCGCGTCTGGTATGAAGA 60

Qy 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40

Db 61 ACCGCTGCTGCTAAATTCGAACGCCAGCACATGCACAGCCAGATCTGGGTACCGACG 120

Qy 41 AspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGly 55

Db 121 GACGACAGGCCATGGCTGATATCGGATCC---GAAGGTGATGTT 162

RESULT 34

AAH43782

ID AAH43782 standard; cDNA; 870 BP.

XX AAH43782;

XX 30-JAN-2002 (first entry)

XX HOP38 coding sequence #5.

XX H. pylori; HOP38; polypeptide delivery system; polar head group;

XX immune stimulating complex; ISCOM; antigenic fragment; saponin;

XX cholesterol; phospholipid; detergent; ss.

XX Helicobacter pylori.

XX WO200176623-A1.

XX 18-OCT-2001.

XX

PF 09-APR-2001; 2001WO-SE00800.

XX 12-APR-2000; 2000GB-0008877.

XX (ASTR) ASTRAZENECA AB.

XX Shapiro A, Sanyal G;

XX WPI; 2002-025884/03.

XX P-PSDB; AAB47732.

XX Production of a polypeptide delivery system useful as a medicament

XX comprises mixing together the polypeptide, cholesterol, saponin, and a

XX phospholipid with a polar head group, in the presence of a detergent

XX Disclosure; Page 45; 48pp; English.

XX The sequences given in AAH43778-82 encode H. pylori HOP38 proteins

XX which were used in the method of the invention. The method of the

XX invention is a process for production of a polypeptide delivery system

XX comprising an immune-stimulating complex (ISCOM) coupled to a

XX polypeptide of H. pylori or its antigenic fragment. The method comprises:

XX mixing the polypeptide, cholesterol, a saponin, and a phospholipid

XX having a polar head group; and removing the detergent from the mixture

XX to form an ISCOM. The method has a broad applicability to polypeptides,

XX including polypeptides that are unsuitable to prior art processes.

XX Sequence 870 BP; 235 A; 170 C; 214 G; 251 T; 0 other;

SQ

Alignment Scores:

Pred. No.: 2,57e-14 Length: 870

Score: 278.50 Matches: 53

Percent Similarity: 96.36% Conservative: 0

Best Local Similarity: 96.36% Mismatches: 1

Query Match: 12.24% Indels: 1

DB: 22 Gaps: 1

US-10-057-531A-2 (1-431) x AAH43782 (1-870)

Qy 1 MethHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20

Db 1 ATGCACCATCATCATCATCATCTCTGGTCTGGTCCACCGCGTCTGGTATGAAGA 60

Qy 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40

Db 61 ACCGCTGCTGCTAAATTCGAACGCCAGCACATGCACAGCCAGATCTGGGTACCGACG 120

Qy 41 AspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGly 55

Db 121 GACGACAGGCCATGGCTGATATCGGATCC---GAAGGTGATGTT 162

RESULT 35

ABA95618

ID ABA95618 standard; cDNA; 2472 BP.

XX ABA95618;

XX 21-MAR-2002 (first entry)

XX Thioredoxin/deoxyxylulose 5-phosphate synthase fusion protein cDNA.

XX DXPS; deoxyxylulose 5-phosphate synthase; pyruvate; thioredoxin;

XX glyceraldehyde 3-phosphate; plant growth modulator;

XX microbial growth modulator; enzyme; gene; ss.

XX Chimeric - Arabidopsis thaliana.

XX Chimeric - Synthetic.

XX Key Location/Qualifiers

XX CDS 1..2472

XX /*tag="a

XX /partial

/product="trxA/txPS fusion protein"
/note="No stop codon given"

US6326164-B1.

04-DEC-2001.

27-JUL-2000; 2000US-0626589.

27-JUL-2000; 2000US-0626589.

(PARA-) PARADIGM GENETICS INC.

Rice JW, Kloti AS, Crawford JM, Lanning B, Stewart SJ;

WPI; 2002-121106/16.

P-PSDB; AN48247.

Assay for determining deoxyxylulose 5-phosphate synthase activity by measuring pyruvate depletion, useful for screening compounds that inhibit or enhance this activity which is useful for modulating plant and microbial growth.

Disclosure: Columns 27-30; 26pp; English.

The present sequence is the coding sequence for a fusion protein comprising a truncated deoxyxylulose 5-phosphate synthase (DXPS) from *Arabidopsis thaliana* and thioredoxin (trxA) from vector pET32. The N-terminal 58 amino acids of the DXPS protein were removed to generate the truncated protein. The truncated DXPS protein possesses DXPS activity. The full-length DXPS is given in AN48245. The present invention relates to a method for determining DXPS activity. The method comprises contacting pyruvate, and optionally glyceraldehyde 3-phosphate with DXPS and then determining the concentration of remaining pyruvate and/or glyceraldehyde 3-phosphate. The method is useful for screening for inhibitors and enhancers of DXPS activity which will have use as modulators of plant and microbial growth.

Sequence 2472 BP; 662 A; 517 C; 640 G; 653 T; 0 other;

Alignment Scores:

Pred. No.:	9,1e-14	Length:	2472
Score:	278.50	Matches:	76
Percent Similarity:	49.75%	Conservative:	25
Best Local Similarity:	37.44%	Mismatches:	49
Query Match:	12.24%	Indels:	53
DB:	24	Gaps:	7

US-10-057-531A-2 (1-431) x ABA95618 (1-2472)

QY 1 MethHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
DB 346 ATGCACCATCATCATCTCTCTGCTGGTCCACGCGGTCTGGTATGAAGAA 405
QY 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40
DB 406 ACCGCTGCTGCTAAATTCGAAGCCACGACATGGACAGCCAGATCTGGGTACCGACGAC 465
QY 41 AspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMetAlaIleSer 60
DB 466 GACGACAGGCCATGGCTGATATCGATCCGCTTCA----- 501
QY 61 ValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLys 80
DB 501 ----- 501
QY 81 ProLeuAla-----GlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePhe 98
DB 502 ---CTTCGACAGAGGGTGAATATTATTCAC---AACAGACCCACCACTCCATTACTTGAC 555
QY 99 ThrPheAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPhe 118
DB 556 ACTATTAACTACCACTCCAC---ATGAAATAATCTTTCTGTCAAGGAAGTGAACAACACTT 612

QY 119 LeuAspValLeuGluSerAspLeuMetGlnPheLys-----HisIle 132
DB 613 TCTGATGAGCTGAGATCAGACGTGATC---TTAATGTGCGAAACCGGTGACATTTG 669
QY 133 SerSerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsn 152
DB 670 GGTCAAGTCTTGGTGTGGAG----- 693
QY 153 ThrLeuLeuLysSerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAla 172
DB 694 ---CTTACTGTGGCTCTTCAATATCTCCACAGACAAGATTTCTTGGGAT 750
QY 173 GlnGluGlyIleSerTyrGluLysValLeuAlaLysTyrLysAspLeuGluSer 192
DB 751 GTTGGTCATCAGTCTTATCTCCTCATAAGATCTTACTGGGAGAGAGAAAGATGCCTACA 810
QY 193 IleLysLys 195
DB 811 ATCAGGCAA 819

RESULT 36

AAA61156
ID AAA61156 standard; DNA; 2079 BP.

XX
AC AAA61156;

XX
DT 12-OCT-2000. (first entry)

XX
DE SEN virus genome fragment SEQ ID NO: 103.

XX
KW SEN virus; SENV; gastrointestinal tract disorder; inflammatory disease; proliferative disorder; hepatopathy; hepatitis; viral infection; vaccination; gene therapy; ds.

XX
OS Hepatitis virus.

XX
PN WO200028039-A2.

XX
PD 18-MAY-2000.

XX
PF 09-NOV-1999; 99WO-EP08566.

XX
PR 10-NOV-1998; 98IT-MI02437.

XX
PR 30-APR-1999; 98IT-MI00923.

XX
PR 14-MAY-1999; 99EP-0830298.

XX
PR 16-JUL-1999; 99EP-0113932.

XX
PA (DIAS-) DIASORIN SRL.

XX
PI Primi D, Fiordalisi G, Mantero GL, Mattioli S, Sottini A;
PI Bonelli F, Vaglini L, Olivero P, Dal Corso A, Bonelli M;

XX
DR WPI; 2000-376551/32.

XX
PT Nucleic acids representing the genome of the SEN virus (SENV) and encoded proteins, useful for treatment of hepatopathies, inflammatory diseases and proliferative disorders such as cancer.

XX
PS Example 16; Page 76-77; 392pp; English.

CC
CC The present invention is concerned with the sequence of the genome of the SEN virus (SENV), and the proteins encoded by it. SENV is thought to be the cause of hepatopathies which are not linked to the presence of the hepatitis A, B and E viruses in man. The genome and proteins of this virus can be used in gene therapy and vaccination against the virus, which also causes disorders of the gastrointestinal tract, including Crohn's disease and lupus erythematosus, inflammatory diseases, and proliferative disorders such as cancer.

XX
SQ Sequence 2079 BP; 656 A; 575 C; 432 G; 416 T; 0 other;

Alignment Scores:

```

Pred. No.: 8,95e-14 Length: 2079
Score: 277.50 Matches: 78
Percent Similarity: 52.00% Conservative: 13
Best Local Similarity: 44.57% Mismatches: 46
Query Match: 12.20% Indels: 39
DB: 21 Gaps: 6

US-10-057-531a-2 (1-431) x AAA61156 (1-2079)
Qy 1 MethHisHisHisHisHisSerGlyLeuValProArgGlySerGlyMetLysGlu 20
Db 1 ATGCACCATCATCATCTCTCTGCTGGTCCACCGGGTCTGGTATGAAGAA 60
Qy 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40
Db 61 ACCGCTGCTGTAATTCGAACGCGACATGGACAGCCAGATCTGGGTACCGAGCAG 120
Qy 41 AspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMetAlaIleSer 60
Db 121 GACGACAAAGCGCTGGCTGATATCGGATCCATGAAC----- 156
Qy 61 ValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLys 80
Db 157 -----TATGCCATGCACTGCCGAGACAGACAGCCGCCAGAA 195
Qy 81 ProLeuAlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIlePheThrPhe 100
Db 196 CCTTTCGGGGC-----GACATGTCACCAATAACTTTC 228
Qy 101 Asn---LeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeu 119
Db 229 AGCCTGCTGGTACTATACGACAGCAGACACCTTAACAGGTGGACCTTCCCAAC 288
Qy 120 AspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIle 139
Db 289 GACCAGCTA-----GACCTGGTGGGTACAAACAC----- 318
Qy 140 GluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSerTyrLys 159
Db 319 ---ACCAGGTTAAATTTTACAGAAGC-----AAAGACACTGA-CTTATAGTTACCTT 368
Qy 160 TyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGlu 174
Db 369 TAACATAAAGCCTCCCATGAATGAACGAGACTACTCTCCCGAA 413

RESULT 37
ID AAA09770 standard; DNA; 7375 BP.
AC AAA09770;
XX
XX
XX 23-JUN-2000 (first entry)
XX
XX
XX Secreted protein acidic and rich in cysteine (SPARC) nucleotide sequence.
XX
XX SPARC; secreted protein acidic and rich in cysteine; pharmaceutical;
XX drug; research reagent; mouse; ds.
XX
XX Mus musculus.
XX
XX JP3012931-B1.
XX
XX 28-FEB-2000.
XX
XX 26-FEB-1999; 99JP-0049826.
XX
XX 26-FEB-1999; 99JP-0049826.
XX
XX (AGEN ) KOGYO GIJUTSUINCHO.
XX
XX WPI; 2000-259560/23.
XX
XX P-PSDB; AAY85150.
XX

PT New secretion protein acidic and rich in cysteine (SPARC) for drugs and
XX research reagents consists of specific amino acid sequence -
XX
XX Claim 3; Page 8-15; 17pp; Japanese.
XX
XX This sequence represents a nucleotide sequence encoding a mouse protein,
XX the protein has the physiological activity of SPARC (secreted protein
XX acidic and rich in cysteine). The protein is used for the preparation of
XX drugs, research reagents and pharmaceuticals. The SPARC protein is
XX prepared efficiently and inexpensively in large quantities. The protein
XX can be used in pharmaceuticals as it does not lose its physiological
XX activity.
XX
XX Sequence 7375 BP; 1760 A; 1955 C; 1945 G; 1715 T; 0 other;
XX

Alignment Scores:
Pred. No.: 6,72e-13 Length: 7375
Score: 275.00 Matches: 50
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 98.04% Mismatches: 0
Query Match: 12.09% Indels: 0
DB: 21 Gaps: 0

US-10-057-531a-2 (1-431) x AAA09770 (1-7375)
Qy 1 MethHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
Db 5554 ATGCACCATCATCATCTCTCTGCTGGTCCACCGGGTCTGGTATGAAGAA 5613
Qy 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40
Db 5614 ACCGCTGCTGTAATTCGAACGCGACATGGACAGCCAGATCTGGGTACCGAGCAG 5673
Qy 41 AspAspLysAlaMetAlaAspIleGlySerIle 51
Db 5674 GACGACAAAGCGCTGGCTGATCGGATCCATG 5706

RESULT 38
AAZ98759
ID AAZ98759 standard; cDNA; 7375 BP.
XX
XX AAZ98759;
XX
XX 20-JUN-2000 (first entry)
XX
XX Murine SPARC nucleotide sequence.
XX
XX Secreted protein acidic and rich in cysteine; SPARC; mouse; nerve cell;
XX neurocyte cell adhesion; cell protrusion retraction; cell shrinkage;
XX nervous system disease; epilepsy; arteriosclerosis; wound healing; ss.
XX
XX Mus sp.
XX
XX JP3012930-B1.
XX
XX 28-FEB-2000.
XX
XX 26-FEB-1999; 99JP-0049708.
XX
XX 26-FEB-1999; 99JP-0049708.
XX
XX (AGEN ) KOGYO GIJUTSUINCHO.
XX
XX WPI; 2000-306484/27.
XX
XX P-PSDB; AAY79067.
XX
XX Drug composition for suppressing neurocyte cell adhesion, generating
XX cell migration and promoting shrinkage retraction of the cell contains
XX specific amino acid sequence
XX
XX Example 2; Page 10-12; 21pp; Japanese.
XX
XX This sequence represents a nucleotide sequence encoding a protein
XX

```

CC referred to as SPARC (secreted protein acidic and rich in cysteine). The
 CC protein is used as the active ingredient in a drug composition for
 CC suppressing neurocyte cell adhesion, generating cell migration and
 CC promoting a nerve protrusion shrinkage reaction. The composition can be
 CC used in wound healing and also to treat diseases of the nervous system,
 CC arteriosclerosis, and epilepsy.

XX SQ Sequence 7375 BP; 1760 A; 1955 C; 1945 G; 1715 T; 0 other;

Alignment Scores:
 Pred. No.: 6,72e-13 Length: 7375
 Score: 275.00 Matches: 50
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 98.04% Mismatches: 0
 Query Match: 12.09% Indels: 0
 DB: 21 Gaps: 0

US-10-057-531A-2 (1-431) x AAZ98759 (1-7375)

QY 1 MetHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
 |||||
 Db 5554 ATGCACCATCATCATCATCTCTCTGGTCTGTCGCCACGGCTTCTGGTATGAAAGAA 5613
 QY 21 ThrAlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40
 |||||
 Db 5614 ACCGCTGCTGCTAAATTCGAACGCCACGACATGACAGCCACAGATCTGGGTACCGACGAC 5673
 QY 41 AspAspLysAlaMetAlaAspIleGlySerIle 51
 |||||
 Db 5674 GAGCACAAAGCCATGCTGCTATATCGATCCATG 5706

RESULT 39

ID AAA94064
 ID AAA94064 standard; DNA; 241 BP.

XX AC AAA94064;

XX DT 30-JAN-2001 (first entry)

DE H. contortus clone 65e vector pET30a cloning junction DNA #2.

XX Nematode; parasite; helminth; sheep; goat; stomach; vaccine; ds.

XX OS Synthetic.

XX FH Key Location/Qualifiers

XX CDS 14..241

XX FT /*tag- a

XX FT /product- "cloning junction sequence"

XX FT /partial

XX PN WO200056763-A1.

XX PD 28-SEP-2000.

XX PF 16-MAR-2000; 2000WO-AU00210.

XX PR 18-MAR-1999; 99AU-0009297.

XX XX (NOVS) NOVARTIS AG.

PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.

XX PI Savin KW, Cook VR, Chen Y, Sexton JL, Apos E, Wilson LR;

XX PI Griffiths TM, Newton SE;

XX DR WPI: 2000-594573/56.

XX DR P-PSDB; AAB26110.

XX PT New Haemonchus contortus polypeptide for inducing a protective effect
 PT against a helminth by controlling helminth infection, growth, viability
 PT and/or egg fecundity and for ameliorating the symptoms of helminth
 PT infection -

XX

PS Disclosure; Fig 8; 94pp; English.

XX The present sequence comprises the sequence found at the cloning
 CC junction in the pET30a vector of nematode Haemonchus contortus
 CC clone 65e. This clone encodes a novel protein. This organism is a
 CC parasite found in the stomach of its host (sheep and goats in
 CC particular). The protein can be used in a vaccine against other
 CC helminths, as well as Haemonchus contortus, including trematodes,
 CC cestodes, nematodes and acanthocephala. These are all capable of causing
 CC severe illness in their hosts, which include sheep, pigs, cattle,
 CC horses, donkeys, dogs, cats, guinea pigs and cage-birds, along with
 CC humans. Antibodies to the protein can be used to diagnose infection.
 CC Note: This sequence is stated as being the same as that shown in SEQ ID
 CC NO: 5 of the specification (see AAA94054). However, this sequence is
 CC shorter than the one shown here.

XX SQ Sequence 241 BP; 69 A; 64 C; 59 G; 49 T; 0 other;

Alignment Scores:

Pred. No.: 1.29e-14 Length: 241
 Score: 274.00 Matches: 50
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 12.04% Indels: 0
 DB: 21 Gaps: 0

US-10-057-531A-2 (1-431) x AAA94064 (1-241)

QY 1 MetHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
 |||||

Db 14 ATGCACCATCATCATCATCTCTCTGGTCTGTCGCCACGGCTTCTGGTATGAAAGAA 73

QY 21 ThrAlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40
 |||||

Db 74 ACCGCTGCTGCTAAATTCGAACGCCACGACATGACAGCCACAGATCTGGGTACCGACGAC 133

QY 41 AspAspLysAlaMetAlaAspIleGlySer 50
 |||||

Db 134 GAGCACAAAGCCATGCTGCTATATCGGATCC 163

RESULT 40

AAAD18937/C

ID AAD18937 standard; DNA; 428 BP.

XX AC AAD18937;

XX DT 18-DEC-2001 (first entry)

XX DE Clone 6 DNA related to the invention.

XX KW Recombinant protein; dietary supplement; therapy: hypertension;
 XX KW hypotensive; ds.
 XX OS Unidentified.

XX PN WO200168674-A2.

XX PD 20-SEP-2001.

XX PF 13-MAR-2001; 2001WO-US07957.

XX PR 13-MAR-2000; 2000US-188990P.

XX PA (MONS) MONSANTO TECHNOLOGY LLC.

XX PI Wang Q, Zhonghong G, Baggot BO, Hadfield K, Zhao J, Edwards J;

XX DR WPI: 2001-607456/69.

XX PT Producing polynucleotides having repeats, by annealing first
 PT polynucleotide pool having tandem repeats of sequences encoding
 PT recombinant protein with its complement pool, extending 3' ends and
 PT denaturing products -

XX Example 4; Fig 7; 77pp; English.

PS
XX
XX
CC The invention relates to the field of recombinant DNA technology
CC and, more particularly, to methods for the production of recombinant
CC polynucleotide sequences and recombinant proteins containing repeating
CC sequences. The invention provides methods for the production of
CC polynucleotides and/or proteins containing tandem repeats of short
CC sequences in which complementary polynucleotides anneal and act as
CC primers allowing extension of their 3' ends to produce long sequences
CC of tandem repeats. The method is useful for producing polynucleotide
CC containing repeating units and for producing recombinant protein
CC containing tandem units. The proteins produced can be used as dietary
CC supplements in the treatment and control of hypertension. The present
CC sequence is a polynucleotide produced by the method of the present
CC invention from Clone 6.

XX
SQ Sequence 428 BP; 85 A; 88 C; 119 G; 127 T; 9 other;

Alignment Scores:

Pred. No.:	2.59e-14	Length:	428
Score:	274.00	Matches:	50
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	12.04%	Indels:	0
DB:	22	Gaps:	0

US-10-057-531A-2 (1-431) x AAD18937 (1-428)

QY	1	MethHisHisHisHisHisSerSerGlyLeuValProAtgGlySerGlyMetLysGlu	20
DB	372	ATGCACCATCATCATCATCTTCTCTGGTCGGCCACCGGTCCTGGTATGAAGAA	313
QY	21	ThrAlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp	40
DB	312	ACCGCTGCTGCTAAATTCGAACGCCAGCACATGGAGACGCCAGATCTGGGTACCGACGAC	253
QY	41	AspAspLysAlaMetAlaAspIleGlySer	50
DB	252	GACGACAAGGGCATGGCTGATATCGGATCC	223

Search completed: May 19, 2003, 15:10:41
Job time : 261 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 19, 2003, 15:04:10 ; Search time 47 Seconds
(without alignments)
2812.291 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 2275

Sequence: 1 MHHHHHHSSGLVPRGSKGME.....TCCTKPDVSLFDGIFCSS 431

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US10057531/runat_12052003_091032_21701/app_query.fasta_1.583
-DB=Issued Patents.NA -OMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10057531 -ECGN_1_1_32_runat_12052003_091032_21701 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Issued Patents.NA:*
- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
 - 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
 - 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
 - 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
 - 5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq:*
 - 6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1935.5	85.1	5181	1	US-08-257-073-10 Sequence 10, Appl
2	1931	84.9	1219	4	US-08-195-705-1 Sequence 1, Appl
3	278.5	12.2	2472	4	US-09-626-589-6 Sequence 6, Appl
4	275	12.1	7375	4	US-09-513-442-1 Sequence 1, Appl
5	274	12.0	501	4	US-09-741-243C-3 Sequence 3, Appl
6	274	12.0	1653	4	US-09-741-243C-1 Sequence 1, Appl
7	259.5	11.4	1227	2	US-08-683-007A-1 Sequence 1, Appl
8	259.5	11.4	6619	2	US-08-683-007A-3 Sequence 3, Appl
9	242	10.6	587	4	US-09-166-966E-1 Sequence 1, Appl
10	173.5	7.6	1146	4	US-09-277-716-21 Sequence 21, Appl
11	173.5	7.6	1146	4	US-09-609-161B-21 Sequence 21, Appl
12	152.5	6.7	867	4	US-09-273-839A-7 Sequence 7, Appl

13	151	6.6	2136	3	US-08-946-475-8 Sequence 8, Appl
14	151	6.6	2136	4	US-09-340-479-8 Sequence 45, Appl
15	148.5	6.5	440	4	US-09-470-191-45 Sequence 8, Appl
16	139.5	6.1	3033	4	US-09-134-001C-2341 Sequence 2341, Ap
17	133	5.8	30549	4	US-09-134-001C-322 Sequence 322, App
18	131	5.8	3763	1	US-07-792-865D-1 Sequence 1, Appl
19	128	5.6	3883	2	US-08-468-036-33 Sequence 33, Appl
20	128	5.6	3883	2	US-08-376-843-33 Sequence 33, Appl
21	128	5.6	3884	4	US-09-541-782-3 Sequence 3, Appl
22	128	5.6	3884	4	US-09-723-820-3 Sequence 3, Appl
23	127.5	5.6	2836	4	US-09-157-257-3 Sequence 3, Appl
24	126.5	5.6	2637	4	US-09-134-001C-1541 Sequence 1541, Ap
25	126.5	5.6	6773	4	US-09-166-350-27 Sequence 27, Appl
26	126	5.5	3466	1	US-08-468-036-38 Sequence 38, Appl
27	126	5.5	3466	2	US-08-376-843-38 Sequence 113, App
28	126	5.5	3628	3	US-08-480-640A-113 Sequence 113, App
29	126	5.5	3628	3	US-08-295-802-113 Sequence 113, App
30	126	5.5	3628	4	US-08-488-237A-113 Sequence 113, App
31	126	5.5	3628	4	US-08-375-992A-113 Sequence 113, App
32	126	5.5	3942	3	US-08-480-640A-189 Sequence 189, App
33	126	5.5	3942	4	US-08-488-237A-189 Sequence 189, App
34	126	5.5	3942	4	US-08-375-992A-189 Sequence 189, App
35	126	5.5	5785	3	US-08-480-640A-221 Sequence 221, App
36	126	5.5	5785	4	US-08-686-968C-221 Sequence 221, App
37	126	5.5	5785	4	US-08-488-237A-221 Sequence 221, App
38	126	5.5	5785	4	US-08-287-959-2 Sequence 2, Appl
39	125.5	5.5	7573	1	US-08-235-836C-73 Sequence 73, Appl
40	125	5.4	2107	4	US-09-134-001C-931 Sequence 931, App
41	123	5.4	8133	1	US-08-480-604A-5 Sequence 5, Appl
42	122.5	5.4	8133	2	US-08-405-496A-5 Sequence 5, Appl
43	122.5	5.4	8133	4	US-08-915-136-5 Sequence 5, Appl
44	122.5	5.4	8133	4	US-08-957-310-5 Sequence 5, Appl
45	122.5	5.4	8133	4	US-08-957-310-5 Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-257-073-10
; Sequence 10, Application US/08257073
; Patent No. 5765597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: NEW YORK
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:

NAME: Frommer, William S.
 REGISTRATION NUMBER: 25,506
 REFERENCE/DOCKET NUMBER: 454310-2570
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 840-3333
 TELEFAX: (212) 840-0712

TELEX: 425066 CURTMS

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 5181 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-257-073-10

Alignment Scores:

Pred. No.: 5,3e-201

Score: 1935.50

Percent Similarity: 96.45%

Best Local Similarity: 94.42%

Query Match: 85.08%

DB: 1

US-10-057-531A-2 (1-431) x US-08-257-073-10 (1-5181)

Qy 38 ThrAspAspLysAlaMetAlaAspLysGlySerLeuGluGlyArgGlyThrMet 57

Db 3952 TCCGAAGATAATGACGAATATTTAGATCAAGTAGTAACTGGAGAA----- 3996

Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77

Db 3997 GCAATATCTGTCACAAATGGGATAATATCTCTCAGGATTTTGAATAATGAATATGATGTTATA 4056

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97

Db 4057 TATTTAAACCTTTAGCTGGAGTATATAGAAGCTTAAACAAACAAATGAAAAACATT 4116

Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117

Db 4117 TTTACATTTAATTTAAATTTGAACGATATCTTAAATTCAGCTCTTAAGAACGAAAAATAT 4176

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerAsnGluTyr 137

Db 4177 TTCTTAGATGTATAGAACTGATTTAATGCAATTTAAACATATATCTCAATGAATAC 4236

Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSer 157

Db 4237 ATTATTCGAAGATTCATTTAAATTTAATTCAGAACAAAAACACACATTTTAAAAAGT 4296

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177

Db 4297 TACAAATATATAAAGAAATCAGTAGAATAATGATATTAATTTGCACAGGAGGTATAGT 4356

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle 197

Db 4357 TATTTAGAAAGGTTTGTAGCAATATATAGGATGATTTAGATCAATTAATAAAGTTATC 4416

Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217

Db 4417 AAAGAAGAAAGGAGAGAGTTCCCATCATCATCACCAACAAACACACCTCGTCACCAAGAAAA 4476

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237

Db 4477 ACAGACCAACAAAGAGAGAGAGTTCCCTCCATTTTAAACAAACATTTGAGACCTTA 4536

Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257

Db 4537 TACATAACTTAGTATATAAATTTGACGATTTACTTAACTTAACGCAAGCAATTAAC 4596

Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277

Db 4597 GATTGTAAATGTTGAAAGAGATGAAGCACATGTTTAAATACTAACTTAACTTAAATA 4656

Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297

Db 4657 GCAATTCATGACAAAATAGATCTTTTTAAAAACCATACGACTTCGAAGCAATTTAAAAA 4716

Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317

Db 4717 TTGATAAATGATGATACGAAAAAGATATGCTTGGCAAAATTTACTTAGTACAGGATAGTT 4776

Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337

Db 4777 CAAATTTTCTTAATACAATAATATCAAAATTAATTAAGAGGAAAAATTTCCCAAGATATGTTA 4836

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357

Db 4837 AACATTTCAACAACACCAATGCGTAAAAAACAATGTCAGAAAAATTTCTGGATGTTTCAGA 4896

Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377

Db 4897 CATTTAGATGAAGAGAGAGATGTAATGTTTATTAATTAACAACAAGAGGTGATAAA 4956

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlnGlyCysAspAlaAspAla 397

Db 4957 TGTGTTGAAAAATCCAAATCTTACTTGTACGAAAAATTAATGTTGGATGTTGATGCGAGATGCC 5016

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417

Db 5017 AAATGTACCGAAGAAGATTCAGGTAGCAACGAAAGAAAAATCACATGTGAATGTACTAAA 5076

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 431

Db 5077 CCGTATCTTATCCACTTTTCGATGGTATTTTCGCAGTTCC 5118

RESULT 2

US-08-195-705-1

; Sequence 1, Application US/08195705

; Patent No. 6420523

; GENERAL INFORMATION:

; APPLICANT: Chang, Sandra

; APPLICANT: Hul, George

; APPLICANT: Barr, Philip

; APPLICANT: Gibson, Helen

; TITLE OF INVENTION: RACULOVIRUS PRODUCED PLASMIDIUM

; TITLE OF INVENTION: FALCIPARUM VACCINE

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Davis Hoxie Faithfull Hapgood

; STREET: 45 Rockefeller Pl.

; CITY: New York

; STATE: N.Y.

; COUNTRY: USA

; ZIP: 10111

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/195,705

; FILING DATE: 14-FEB-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Jacobs, Seth H

; REGISTRATION NUMBER: 32140

; REFERENCE/DOCKET NUMBER: 11880A3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-757-2200

; TELEFAX: 212-586-1461

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1219 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium falciparum
 STRAIN: falciparum uganda palo alto
 IMMEDIATE SOURCE:
 CLONE: f195LFP42acNPV
 FEATURE:
 NAME/KEY: exon
 LOCATION: 13..1194
 US-08-195-705-1

Alignment Scores:
 Pred. No.: 1,78e-201 Length: 1,219
 Score: 1931.00 Matches: 369
 Percent Similarity: 98.67% Conservative: 1
 Best Local Similarity: 98.40% Mismatches: 5
 Query Match: 84.88% Indels: 0
 DB: 4 Gaps: 0

US-10-057-531A-2 (1-431) x US-08-195-705-1 (1-1219)

Qy	56	ThrMetAlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAsp	75
Db	70	ACCGGCGGATATCTGTCCACAAATGATATATCTCTCAGGATTTGAAATGAATGAT	129
Qy	76	ValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIleGluLys	95
Db	130	GTTATATATTAAACCCCTTACGTGGAGTATATAGACCTTAAACAAACAAATTTGAAAA	189
Qy	96	AsnIlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArg	115
Db	190	AACATTTTACATTTAAATTTAAATTTGAACGATATCTTAAATTCACGCTTAAAGAACCA	249
Qy	116	LysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsn	135
Db	250	AAATATTCTTATGATGATTAAGAATCTGATTAATGCAATTTAAACATATATCTCTCAAT	309
Qy	136	GluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeu	155
Db	310	GAATACATATTGAAGATTCATTTAAATTTGAATTTGAGAACAAAAAACACATTTTA	369
Qy	156	LysSerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGly	175
Db	370	AAAAAGTTACAAATATATAAAGAAATCAGTACGAAATGATATTAATTTGCACAGGAAGT	429
Qy	176	IleSerTyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLys	195
Db	430	ATAAGTTATTATGAAAGGTTTACGCAAAATATAAGGATGATTTAGAATCAATTAATAAA	489
Qy	196	ValIleLysGluGluLysGluLysPheProSerProProThrProProSerPro	215
Db	490	GTTATCAAGAAGAAAGGAGAGTTCCCATCATCACCACCAACACACCTCCGTCACCA	549
Qy	216	AlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGlu	235
Db	550	GCAAAACAGACGACAAACAAAGAAAGTAAGTTCCCTTCCATTTTAAACAAACATGAG	609
Qy	236	ThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLys	255
Db	610	ACCTTATACATAACTTAGTTAAATAAATTTGACGATTACTTAATTAACCTTAAAGGCAAG	669
Qy	256	IleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAsp	275
Db	670	ATTAAACGATTTGAATTTGAAAAAGATGAACACATGTTAAATAACTTAACCTTAGTCAT	729
Qy	276	LeuLysAlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIle	295
Db	730	TTAAAGCAATTTGATGACAAATAGATCTTTTAAAAAACCAATACCACTTCGAACCAAT	789
Qy	296	LysLysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGly	315

Qy 41 AspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMetAlaIleSer 60
Db 466 GACGACAAAGGCGATGCTGATACCGCTTCA-----501
Qy 61 ValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyraSpValIleTyLeuLys 80
Db 501 -----501
Qy 81 ProLeuAla-----GlyValTyraArgSerLeuLysLysGlnIleGluLysAsnIlePhe 98
Db 502 ---CTTGACAGAGCGGTGAATATTATTC---AACAGACCACCACTCCATTACTTGCAC 555
Qy 99 ThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrrPhe 118
Db 556 ACTATTAACCTACCCCAATCCAC---ATGAAATCTTCTGTCAGGAAGAACTGAAACAACATT 612
Qy 119 LeuAspValLeuGluSerAspLeuMetGlnPheLys-----HisIle 132
Db 613 TCTGATGAGCTGAGTACAGAGTGATC---TTTATGTGTCMAAACCGGTGGACATTIG 669
Qy 133 SerSerAsnGluTyrrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsn 152
Db 670 GGGTCAAGCTTGGTGTGGAG-----693
Qy 153 ThrLeuLeuLysSerTyrrLysTyrrIleLysGluSerValGluAsnAspIleLysPheAla 172
Db 694 ---CTTACTGGCTTCCTCATTACATTTCAATCTCCACAGAACAAAGATTCTTTGGGAT 750
Qy 173 GlnGluGlyIleSerTyrrTyrrGluLysValLeuAlaLysTyrrLysAspAspLeuGluSer 192
Db 751 GTTGTGCATCAGCTTATCCTCATAAGATTCTTACTGGGAGAGAGAAAGATGCTTACA 810
Qy 193 IleLysLys 195
Db 811 ATGAGGCAA 819
RESULT 4
US-09-513-442-1
; Sequence 1, Application US/09513442
; Patent No. 6387664
; GENERAL INFORMATION:
; APPLICANT: Ikemoto, Mitsushi
; TITLE OF INVENTION: SPARC Fusion Protein and a Process for Producing the
; TITLE OF INVENTION: Same
; FILE REFERENCE: HIRAKI-04218
; CURRENT APPLICATION NUMBER: US/09/513.442
; PRIOR FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 7375
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5209)..(6609)
US-09-513-442-1
Alignment Scores:
Pred. No.: 2.15e-19 Length: 7375
Score: 275.00 Matches: 50
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 98.04% Mismatches: 0
Query Match: 12.09% Indels: 0
DB: 4 Gaps: 0
US-10-057-531A-2 (1²431) x US-09-513-442-1 (1-7375)
Qy 1 MethHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
Db 5554 ATGCACCATCATCATCATCTTCTGCTGCTGCCACGCGTCTCTGGTATGAAGAA 5613
Qy 21 ThrAlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40

Db 5614 ACCGCTGCTGCTAAATTCGACCGCAGCACATGGACGCCAGATCTGGGTACCGACGAC 5673
Qy 41 AspAspLysAlaMetAlaAspIleGlySerIle 51
Db 5674 GACGACAAAGGCGATGCTGATATCGATCCATG 5706
RESULT 5
US-09-741-243C-3
; Sequence 3, Application US/09741243C
; Patent No. 6399352
; GENERAL INFORMATION:
; APPLICANT: Crawford Jr., John Milton
; APPLICANT: Rice, John
; APPLICANT: Sevala, Veeresh
; APPLICANT: Stewart, Sandy
; TITLE OF INVENTION: NOVEL PLANT PORPHOBILINOGEN AND FUSION
; TITLE OF INVENTION: PROTEIN THEREOF
; FILE REFERENCE: 2022US
; CURRENT APPLICATION NUMBER: US/09/741.243C
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/171,785
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: DNA encoding thioredoxin functional fragment
US-09-741-243C-3
Alignment Scores:
Pred. No.: 4.44e-21 Length: 501
Score: 274.00 Matches: 50
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.04% Indels: 0
DB: 4 Gaps: 0
US-10-057-531A-2 (1²431) x US-09-741-243C-3 (1-501)
Qy 1 MethHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
Db 346 ATGCACCATCATCATCATCTTCTGCTGCTGCCACGCGTCTCTGGTATGAAGAA 405
Qy 21 ThrAlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40
Db 406 ACCGCTGCTGCTAAATTCGACCGCAGCACATGGACGCCAGATCTGGGTACCGACGAC 465
Qy 41 AspAspLysAlaMetAlaAspIleGlySer 50
Db 466 GACGACAAAGGCGATGCTGATATCGATCCATG 495
RESULT 6
US-09-741-243C-1
; Sequence 1, Application US/09741243C
; Patent No. 6399352
; GENERAL INFORMATION:
; APPLICANT: Crawford Jr., John Milton
; APPLICANT: Rice, John
; APPLICANT: Sevala, Veeresh
; APPLICANT: Stewart, Sandy
; TITLE OF INVENTION: NOVEL PLANT PORPHOBILINOGEN AND FUSION
; TITLE OF INVENTION: PROTEIN THEREOF
; FILE REFERENCE: 2022US
; CURRENT APPLICATION NUMBER: US/09/741.243C
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/171,785
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0

```
; SEQ ID NO 1
; LENGTH: 1653
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DNA encoding a plant thioresodoxin-porphobilinogen
; OTHER INFORMATION: synthase fusion protein
US-09-741-243C-1

Alignment Scores:
Pred. No.: 2,78e-20 Length: 1653
Score: 274.00 Matches: 50
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.04% Indels: 0
DB: 4 Gaps: 0

US-10-057-531A-2 (1-431) x US-09-741-243C-1 (1-1653)
QY 1 MethHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
DB 346 ATGCACCATCATCATCATCTCTCTGGTCTGGTCCACGCGGTTCTGGTATGAAGAA 405
QY 21 ThrAlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAsp 40
DB 406 ACCGCTGCTCTAAATTCGAAGCCACGACATGGACAGCCACATCTGGGTACCGACGAC 465
QY 41 AspAspLysAlaMetAlaAspLysGlySer 50
DB 466 GAGCACAAGGCCATGCTGATATCGGATCC 495
QY 61 ValThrMetAspAsnLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLys 80
DB 495 -----ACTGGCCATCGCAGACTAC----- 515
QY 81 ProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPhe 100
DB 516 -----AGGTAGACATATATCT 533
QY 101 AsnLeuAsnLeuAsnAspLysLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAsp 120
DB 534 AACTTGGAGTCAACGAA-----TTTCAAGAC 560
QY 121 -ValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleLeuG 140
DB 561 AATTCTGGAGTGGGAACCCCAATCCATCGATCATGTC-----TACACAGTTCA 608
QY 140 uAspSerPheLysLeuAsnSerGluGlnLysAsnThrLeuLysSerTyrLysTy 160
DB 609 GATAAGCACTAGGCTAGAAACTGGAGAGCAAAATGTTCTTAAACCCGAGACGGAGTG 668
QY 160 r---lleLysGluSerValGluAsnAspLysPheAlaGlnGluGlyIleSerTy 179
DB 669 CGACCTCACCGATGAGTCTGTAAGGACGTG-----GGCGACAGGTACAT 713
QY 179 rGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysValIleLysG 199
DB 714 GGCGCGGGTCTCTCC-----TACCCGCAAGGAACGG 746
QY 199 uGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLysThrAs 219
DB 747 AAACACCACCGGGTTCCTCCGAGGAGCTCTCTTT----- 780
QY 219 pGluGlnLysLysGluSerLysPheLeuProPheLeu---ThrAsnIle 234
DB 781 -----AGAAACTCCCGGAGTTTCACGCCGTACCTGGACACAAACCTC 822

RESULT 8
US-08-683-007A-3/c
; Sequence 3, Application US/08683007A
; Patent No. 5858724
; GENERAL INFORMATION:
; APPLICANT: No. 5858724y, Robert E
; APPLICANT: Domanioco, Michael
; APPLICANT: Yaeger, Keith
; APPLICANT: Kroeker, Warren
```

```
; DESCRIPTION: trxA-rabbit tissue factor fusion protein"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1227
US-08-683-007A-1

Alignment Scores:
Pred. No.: 6,75e-19 Length: 1227
Score: 259.50 Matches: 80
Percent Similarity: 43.04% Conservatives: 22
Best Local Similarity: 33.76% Mismatches: 55
Query Match: 11.41% Indels: 81
DB: 2 Gaps: 9

US-10-057-531A-2 (1-431) x US-08-683-007A-1 (1-1227)
QY 1 MethHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
DB 346 ATGCACCATCATCATCATCTCTCTGGTCTGGTCCACGCGGTTCTGGTATGAAGAA 405
QY 21 ThrAlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAsp 40
DB 406 ACCGCTGCTCTAAATTCGAAGCCACGACATGGACAGCCACATCTGGGTACCGACGAC 465
QY 41 AspAspLysAlaMetAlaAspLysGlySerIleGluGlyArgGlyThrMetAlaIleSer 60
DB 466 GAGCACAAGGCCATGCTGATATCGGATCC----- 494
QY 61 ValThrMetAspAsnLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLys 80
DB 495 -----ACTGGCCATCGCAGACTAC----- 515
QY 81 ProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPhe 100
DB 516 -----AGGTAGACATATATCT 533
QY 101 AsnLeuAsnLeuAsnAspLysLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAsp 120
DB 534 AACTTGGAGTCAACGAA-----TTTCAAGAC 560
QY 121 -ValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleLeuG 140
DB 561 AATTCTGGAGTGGGAACCCCAATCCATCGATCATGTC-----TACACAGTTCA 608
QY 140 uAspSerPheLysLeuAsnSerGluGlnLysAsnThrLeuLysSerTyrLysTy 160
DB 609 GATAAGCACTAGGCTAGAAACTGGAGAGCAAAATGTTCTTAAACCCGAGACGGAGTG 668
QY 160 r---lleLysGluSerValGluAsnAspLysPheAlaGlnGluGlyIleSerTy 179
DB 669 CGACCTCACCGATGAGTCTGTAAGGACGTG-----GGCGACAGGTACAT 713
QY 179 rGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysValIleLysG 199
DB 714 GGCGCGGGTCTCTCC-----TACCCGCAAGGAACGG 746
QY 199 uGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLysThrAs 219
DB 747 AAACACCACCGGGTTCCTCCGAGGAGCTCTCTTT----- 780
QY 219 pGluGlnLysLysGluSerLysPheLeuProPheLeu---ThrAsnIle 234
DB 781 -----AGAAACTCCCGGAGTTTCACGCCGTACCTGGACACAAACCTC 822

RESULT 8
US-08-683-007A-3/c
; Sequence 3, Application US/08683007A
; Patent No. 5858724
; GENERAL INFORMATION:
; APPLICANT: No. 5858724y, Robert E
; APPLICANT: Domanioco, Michael
; APPLICANT: Yaeger, Keith
; APPLICANT: Kroeker, Warren
```



```
; Sequence 21, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; EARLIER FILING DATE: 1999-03-26
; EARLIER FILING DATE: 1998-10-01
; EARLIER FILING DATE: 1998-10-01
; EARLIER FILING DATE: 1998-06-15
; EARLIER FILING DATE: 1998-06-15
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 1146
; TYPE: DNA
; ORGANISM: Artificial Sequence: fusion construct
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1146)
; FEATURE:
; OTHER INFORMATION: DNA encoding a Cellulose Binding Domain (CBD)-Gaussia luciferase
US-09-277-716-21

Alignment Scores:
Pred. No.:          Length:      1146
Score:              173.50      Matches:      35
Percent Similarity: 92.31%      Conservative: 1
Best Local Similarity: 89.74%      Mismatches: 0
Query Match:        7.63%      Indels:      3
DB:                  4          Gaps:          1

US-10-057-531A-2 (1-431) x US-09-277-716-21 (1-1146)
;
; TYPE: DNA
; ORGANISM: Artificial Sequence: fusion construct
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1146)
; FEATURE:
; OTHER INFORMATION: DNA encoding a Cellulose Binding Domain (CBD)-Gaussia luciferase
US-09-277-716-21

Alignment Scores:
Pred. No.:          Length:      1146
Score:              173.50      Matches:      35
Percent Similarity: 92.31%      Conservative: 1
Best Local Similarity: 89.74%      Mismatches: 0
Query Match:        7.63%      Indels:      3
DB:                  4          Gaps:          1

US-10-057-531A-2 (1-431) x US-09-277-716-21 (1-1146)
QY 8 SerSerGlyLeuValProArgGlySer-----GlyMetLysGluThrAlaAla 24
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 475 TCCGGGGTCTGGTGCACGCGGTACTGCAATTGGTATGAAGAAACCGCTGCTGCT 534

QY 25 LysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAspAspLys 43
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 535 AAATTCGAACGCCAGCACATGGACGCCAGATCTGGTACCGATGACGACGACAAG 591

RESULT 11
US-09-609-161B-21
; Sequence 21, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HI
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; PRIOR FILING DATE: 2000-06-30
; PRIOR FILING DATE: 1999-03-26
; PRIOR FILING DATE: 1998-10-01
; PRIOR FILING DATE: 1998-10-01
; PRIOR FILING DATE: 1998-06-15
; PRIOR FILING DATE: 1998-06-15
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 1146
; TYPE: DNA
```

```
; ORGANISM: Artificial Sequence: fusion construct
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1146)
; OTHER INFORMATION: DNA encoding a Cellulose Binding Domain (CBD)-Gaussia lucifera
US-09-609-161B-21

Alignment Scores:
Pred. No.:          Length:      1146
Score:              173.50      Matches:      35
Percent Similarity: 92.31%      Conservative: 1
Best Local Similarity: 89.74%      Mismatches: 0
Query Match:        7.63%      Indels:      3
DB:                  4          Gaps:          1

US-10-057-531A-2 (1-431) x US-09-609-161B-21 (1-1146)
QY 8 SerSerGlyLeuValProArgGlySer-----GlyMetLysGluThrAlaAla 24
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 475 TCCGGGGTCTGGTGCACGCGGTACTGCAATTGGTATGAAGAAACCGCTGCTGCT 534

QY 25 LysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAspAspLys 43
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 535 AAATTCGAACGCCAGCACATGGACGCCAGATCTGGTACCGATGACGACGACAAG 591

RESULT 12
US-09-273-839A-7
; Sequence 7, Application US/09273839A
; Patent No. 6329156
; GENERAL INFORMATION:
; APPLICANT: Cirino, Nick M
; APPLICANT: Jackson, Paul J
; APPLICANT: Lehnert, Bruce E
; TITLE OF INVENTION: Disruption of Anthrax Toxin Binding to Cell Surface
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: S-89,662
; CURRENT APPLICATION NUMBER: US/09/273,839A
; CURRENT FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Bacillus anthracis
US-09-273-839A-7

Alignment Scores:
Pred. No.:          Length:      867
Score:              152.50      Matches:      74
Percent Similarity: 40.88%      Conservative: 65
Best Local Similarity: 21.76%      Mismatches: 96
Query Match:        6.70%      Indels:      105
DB:                  4          Gaps:          19

US-10-057-531A-2 (1-431) x US-09-273-839A-7 (1-867)
QY 2 HisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu--- 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 13 CATCATCATCATCATCACAGCGGCTGGTGGCCGCGGCGCATATGCTCGAGGCG 72

QY 21 ---ThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAsp 39
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 73 GCGCATGCCGCCGAAACACTGCACGTATCATTTTAAATGGAAGAAATTAATCTG--- 129

QY 40 AspAspAspLysAlaMetAlaAspLleGlySerLleGluGlyArgGlyThrMetAla 59
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 ---GTAGAAGCGGATAGCGGGTAAATCTCTAGTATGATCCATTAGAAACGACTAAACCG 186

QY 60 SerValThrMetAspAsnIleLeuSer-----GlyPheGlu----- 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 187 GATATGACATTAAAGAACGCCCTTAAATAGCATTTGGATTTAAACGAACCGAATGGAAC 246
```

```
Qy 72 -----AsnGluTyrAspValIleTyrLeuLysProLeuAla 83
Db 247 TTACAATATCAAGGAAAGACATAACCGAATTTGATTTTAAATTCGATCAACAACATCT 306
Qy 84 GlyValTyrArgSerLeuLysGlnIleGluLys -----AsnIlePheThr 99
Db 307 -----CAAAATATCAAGAACAGTTCAGCGGAATTAACGCACTAATATATAT 417
Qy 100 Phe -----AsnLeuAsnAspIleLeuAsnSerArgLeuLysArgLysThr 117
Db 358 GTATTAGATAAAATCAAAATTAATGCAAAATGAATATTTAATAAGAGATAAACGTTT 417
Qy 118 PheLeuaspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
Db 418 CATTATGAT-----AGAAATAAATAGCA-----GTTGGCGGATGAGTCA 459
Qy 138 IleIleGluAspSerPheLys---LeuLeuAsnSerGluGlnLysAsnThrLeuLys 156
Db 460 GTAGTTAAGGAGGCTCATAGAGAAGTAATTAATTCGTCACAGAGGATTAATGTTA--- 516
Qy 157 SerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIle 176
Db 517 -----AATATTGATAAGGATATAAGA----- 537
Qy 177 SerTyrTyrGluLysValLeuAlaLysTyr-----LysAspAspLeuGluSerIle 193
Db 538 -----AAATATTATCAGGTTATATCTGAGAAATGAAGATCTGAAGGCTT 585
Qy 194 LysLysValIleLysGluGlnLysGluLysPheProSerSerProThrProPro 213
Db 586 AAAGAAGTTAATGATGACAGATATGATGTTGATATTCT----- 627
Qy 214 SerProAlaLysThrAspGluGlnLysLysGluSerLys---PheLeuProPheLeuThr 232
Db 628 -----AGTTTACGCGAAGATGCAAAACATTATATAGATTTT----- 663
Qy 233 AsnIleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspPtyrLeuIleAsnLeu 252
Db 664 -----AAAAATAATAT-----GATAAATTACGTTATATATAANGTAATCCC 705
Qy 253 LysAlaLysIleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLys 272
Db 706 AATTATAAGGTAATGATGATGCTGTACTAAACAAACACTATATT----- 753
Qy 273 LeuSerAspLeuLysAlaIleAspLysIleAspLeuPheLysAsnProTyrAspPhe 292
Db 754 -----AATCCTAGTGAG----- 765
Qy 293 GluAlaIleLysLysLeuIleAsnAspThrLysLysAspMetLeuGlyLysLeuLeu 312
Db 766 -----AATGGGATAGTACTAGTACCACCGGATCAAGAAATTTTA 804

RESULT 13
US-08-946-475-8
; Sequence 8, Application US/08946475
; Patent No. 6013505
; GENERAL INFORMATION:
; APPLICANT: Gwynn, Michael
; APPLICANT: Kallender, Howard
; APPLICANT: Palmer, Leslie
; TITLE OF INVENTION: TOPOISOMERASE I
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
```

```
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,475
FILING DATE: 08-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/027,973
FILING DATE: 08-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50560
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2136 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-946-475-8

Alignment Scores:
Pred. No.: 1,12e-06 Length: 2136
Score: 151.00 Matches: 68
Percent Similarity: 36.03% Conservative: 39
Best Local Similarity: 22.90% Mismatches: 80
Query Match: 6.64% Indels: 110
DB: 3 Gaps: 13

US-10-057-531A-2 (1-431) x US-08-946-475-8 (1-2136)
Qy 2 HisHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMet----- 18
Db 13 CATCATCATCATCATCATCAGCAGCGCTGGTCCGCGCGCATATATGACCTTGGCA 72
Qy 19 -----LysGluThrAlaAlaLysPheGluArgGlnHisMetAsp 32
Db 73 GATAATTTAGTCATGTTGTAATCCCTGCACAAACCAACCATTTGAAGATATTAGGT 132
Qy 33 SerProAspLeuGlyThrAspAspLysAlaMetAlaAspIleGlySerIleGlu 52
Db 133 AAG-----AAATATAAGTTATAGCTTCAATGGGACACAGTCAGA 171
Qy 53 Gly-----ArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSerGlyPhe 70
Db 172 GACTTACCAGAGTCAATGGGTGCGAC---ACTGAAGATAAT-----TAC 216
Qy 71 GluAsnGluTyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLys 90
Db 217 GAACCAAAATAT-----ATAACAATACCGGAAAGGTCCTGTTGTAAGAAGATTGAAA 270
Qy 91 LysGlnIleGlu-----LysAsnIlePheThrPheAsn----- 101
Db 271 AAACATCAAAAAAAGCGGAAAGAGCTCTTCTCGCAAGTGACCCCGACCGGTGAAGGTGAA 330
Qy 102 -----LeuAsnLeuAsnAspIleLeuAsnSerArgLeu 112
Db 331 GCATTCCTTGGCATTTATCAAAATTTTAGAGCTTGAAGATTTCAAGAAATTCGCGTT 390
Qy 113 LysLysArgLysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIle 132
Db 391 GTTTTCAACGAAATACTAAAGAGCGCTGTTAAAGAAAGTTTAAAGAAATTCCTAGAGAATT 450
Qy 133 SerSerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsn 152
Db 451 GAAATGAAC-----TTAGTCGATGCACAAACAGCGGT 483
Qy 153 ThrLeuLeuLysSer-----TyrLysTyrIleLys 162
```

Db 484 CGAATATTAGATAGATTGGTTGGCTATACATCTCGCCAGTTCTATGTGAAAAAAGTGAA 543
Qy 163 GluSerVal-----Glu 166
Db 544 AAAGTTGTCAGCGGTCGAGTTCAATCTGTTGCATCTCGTTAGTATTGACCGTGAA 603
Qy 167 AsnAspIle----- 169
Db 604 AATGAATTCGAACCTTTAAACGAGAAGAAATATTGGACTATTGAAGGAGAATTTAGATAC 663
Qy 170 -----LysPheAlaGlnGluGlyIleSerTyr----- 178
Db 664 AAAAAATCAAAATTCAAATCTAAATTCCTTATATAAAAAATAAACCTTTTAAATTAATA 723
Qy 179 -----TyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLys 194
Db 724 ACGAAAAAGATGTTGAGAAATTTACAGCTGCACATAGATGGAGATCAATTCGAATTTACA 783
Qy 195 LysValIleLysGluLysGluLysPheProSerSerProProThrThr 211
Db 784 AACGTGACTAAAAAGAAAAACGGCTAATCCAGCAAAACCCATTTACAAC 834

RESULT 14

US-09-340-479-8
; Sequence 8, Application US/09340479
; Patent No. 6274139
; GENERAL INFORMATION:
; APPLICANT: Gwynn, Michael
; APPLICANT: Kallender, Howard
; APPLICANT: Palmer, Leslie
; TITLE OF INVENTION: TOPOISOMERASE I
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/340,479
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,475
; FILING DATE: 08-OCT-1997
; APPLICATION NUMBER: 60/027,973
; FILING DATE: 08-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Glimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50560
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-09-340-479-8

Alignment Scores: 1.12e-06 Length: 2136
Pred. No.: 151.00 Matches: 68
Score:

Percent Similarity: 36.03% Conservative: 39
Best Local Similarity: 22.90% Mismatches: 80
Query Match: 6.64% Indels: 110
DB: 4 Gaps: 13
US-10-057-531a-2 (1-431) x US-09-340-479-8 (1-2136)
Qy 2 HisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMet----- 18
Db 13 CATCATCATCATCATCATCAGCAGCGCCCTGGTGGCGGCGGAGCCATATGACCTTGGCA 72
Qy 19 -----LysGluThrAlaAlaLysPheGluArgGlnHisMetAsp 32
Db 73 GATAATTTAGTCATCTTGAATCGCTGCAAAAGCAAAACCATTTGAAAGATATTAGGT 132
Qy 33 SerProAspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGlySerIleGlu 52
Db 133 AG-----AAATATAAGTTTATAGCTTCAATGGGACACGTCAGA 171
Qy 53 Gly-----ArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSerGlyPhe 70
Db 172 GACTTACCAAGAAGTCAATGGGTGCAC---ACTGAAGATAAT-----TAC 216
Qy 71 GluAsnGluTyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLys 90
Db 217 GAACCAAAATAT-----ATAACAATACGCGGAAAAAGTCTCTGTGTAAGAAGAAATTGAA 270
Qy 91 LysGlnIleGlu-----LysAsnIlePheThrPheAsn----- 101
Db 271 AAACATGCAAAAAAGCGAAAAACGCTTTTCGCAAGATGACCCCGACCGTGAAGGTGAA 330
Qy 102 -----LeuAsnLeuAsnAspIleLeuAsnSerArgLeu 112
Db 331 GCAATTGCTTGGCATTATCAAAATTTTAGAGCTTTGAAGATCTTAAAGAAATTCGCGT 390
Qy 113 LysLysArgLysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIle 132
Db 391 GTTTTCAACGAAATAACTAAAGACGCTGTAAAGAAAGTTTAAAAATCTCAGAGAAATT 450
Qy 133 SerSerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsn 152
Db 451 GAATGAAC-----TTAGTCGATGCACAAACGCGCGT 483
Qy 153 ThrLeuLeuLysSer-----TyrLysTyrIleLys 162
Db 484 CGAATATTAGATAGATTGGTTGGCTATACATCTCGCCAGTTCTATGGAAGAAAGTGAA 543
Qy 163 GluSerVal-----Glu 166
Db 544 AAAGGTTTGTACGCGGTCGAGTTCAATCTGTTGCACCTTCGTTTAGTCATTGACCGTGA 603
Qy 167 AsnAspIle----- 169
Db 604 AATGAATTCGAACCTTTAAACGAGAAGAAATATTGGACTATTGAAGGAGAATTTAGATAC 663
Qy 170 -----LysPheAlaGlnGluGlyIleSerTyr----- 178
Db 664 AAAAAATCAAAATTCAAATCTAAATTCCTTCAATATAAAAAATAAACCTTTTAAATTAATA 723
Qy 179 -----TyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLys 194
Db 724 ACGAAAAAGATGTTGAGAAATTTACAGCTGCACATAGATGGAGATCAATTCGAATTTACA 783
Qy 195 LysValIleLysGluLysGluLysPheProSerSerProProThrThr 211
Db 784 AACGTGACTAAAAAGAAAAACGGCTAATTCAGCAAAACCCATTTACAAC 834

RESULT 15

US-09-470-191-45/c
; Sequence 45, Application US/09470191
; Patent No. 6465633
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir

; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Compositions and Methods of Their Use in
 ; FILE REFERENCE: the Treatment, Prevention and Diagnosis of Tuberculosis
 ; CURRENT APPLICATION NUMBER: US/09/470,191
 ; CURRENT FILING DATE: 1999-12-23
 ; PRIOR APPLICATION NUMBER: US 60/113,952
 ; PRIOR FILING DATE: 1998-12-24
 ; NUMBER OF SEQ ID NOS: 97
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 45
 ; LENGTH: 440
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium tuberculosis
 ; FEATURE:
 ; NAME/KEY: modified base
 ; LOCATION: (1)...(440)
 ; OTHER INFORMATION: n - any nucleotide
 US-09-470-191-45

Alignment Scores:
 Pred. No.: 1,86e-07 Length: 440
 Score: 148.50 Matches: 43
 Percent Similarity: 61.33% Conservative: 3
 Best Local Similarity: 57.33% Mismatches: 13
 Query Match: 6.53% Indels: 17
 DB: 4 Gaps: 3

US-10-057-531A-2 (1-431) x US-09-470-191-45 (1-440)

QY 1 MethHisHisHisHisHisHisHisHisSerSerGlyLeuValProArgGly-SerGlyMetLysGI 20
 DB 354 ATGCACCATCATTCATCATCTCTCTGCTGGGCGCANGCGGTTTGTGTATGAAGA 295
 QY 20 uThAla-AlaAlaLysPheGluArgGlnHisMetAspSerPro----- 34
 DB 294 AACCGCTGCTGCTAAATTCGNACGCCGCCANATGACAG-CCGATNCGCATTTAGGG 236
 QY 35 -----AspLeuGlyThrAspAspAspLysAlaMetala-----A 47
 DB 235 ACATATAGNATACCAAGATCTGGGTACCGACGACGACGACGACGATATNGG 176
 QY 47 spileGlySerIleGluGly---ArgGlyThrMetAlaIle 59
 DB 175 ATCCGAATTCGGCAGCGCTTCGTCGACCACTNGGATT 135

RESULT 16
 US-09-134-001C-2341
 ; Sequence 2341, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 2341
 ; LENGTH: 3033
 ; TYPE: DNA
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-2341

Alignment Scores:
 Pred. No.: 3.47e-05 Length: 3033
 Score: 139.50 Matches: 71
 Percent Similarity: 40.31% Conservative: 58
 Best Local Similarity: 22.19% Mismatches: 112

Query Match: 6.13% Indels: 79
 DB: 4 Gaps: 14
 US-10-057-531A-2 (1-431) x US-09-134-001C-2341 (1-3033)
 QY 86 TyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeuAsnLeuAsn 105
 DB 1042 TATCACAATGAATTAAGGTTTTTCAAAACAACTCGCAACATTTATCAACACGAGAAAT 1101
 QY 106 Aspile-----LeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeuGlu 123
 DB 1102 GAATAACTCAATTAATCAGTATCTAGAAAAAACCAAGTTTCTTCAATCAATAGAT 1161
 QY 124 SerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr-----IleIleGluAsp 141
 DB 1162 -----AAGATTATTAGTAGTTATCAACAAAAACCGGTAATTGAAGAA 1203
 QY 142 SerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIle 161
 DB 1204 GAAATAAAAAAGATTATACAGTGAATATAATGATTAAACCAAAAAAGAAAGATTGACG 1263
 QY 162 LysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLys 181
 DB 1264 AAGAATGAACAACAACAAGATTGTTGCA----- 1296
 QY 182 ValLeuAlaLysTyrLysAspAspLysValIleLysLysValIleLysGluGluLys 201
 DB 1297 ATTATTGAACATTACACTGAAGAGATTATTAAGCTGAAAGATTATATAGATGAA----- 1350
 QY 202 GluLysPheProSerSerProThrProSerProAlaLysThrAspGluGln 221
 DB 1351 -----TCTGAAGACAA 1362
 QY 222 LysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyr----- 238
 DB 1363 AAAAGGATGAGAAATTTTGTATAATTCAACTAGATAAATCATCTTATCTAGCAAA 1422
 QY 239 -----AsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAla 254
 DB 1423 TTAAGAGAGAAAGAACAGTTAAATGAATTAATGAATCAATCAATCAATATATAGATCG 1482
 QY 255 LysIleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSer 274
 DB 1483 ACTTAATTGATTGAATGACAAAAAGGAT-----TTTGTAAATGAATAAAG 1530
 QY 275 AspLeuLysAlaIleAspAspLysIleAspLeuPheLysAsnPro----- 289
 DB 1531 TCCGCTATGTCAATTGGAGATACCTGTCCAATTTGTGTAATGAATATACATTCATTGGGA 1590
 QY 290 -----TyrAspPheGluAlaIleLysLysLeuIleAsnAspAspThrLysLysAspMet 307
 DB 1591 GAACATATTGATTGTAATCAATTGCTCAA-----AAAAATAATAAAA 1632
 QY 308 LeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThrIleIleSerLys 327
 DB 1633 ATAAACCGTTAGAAAGTACAGAGGTA-----AAATTCGTGATGAATAAATCAAA--- 1683
 QY 328 LeuIleGluGlyLysPheGlnAspMet-----LeuAsnIleSerGln 341
 DB 1684 ---ATAGAACTCGAATTGAAGAAATTAATCATAGAGAAATAATGAATTTTGAAGAA 1740
 QY 342 HisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGlu 361
 DB 1741 CAAGAA-----AAGAAGGATATTAGTGAGCTTCAAAAAACAGCTTAATCATTTGAATCAA 1794
 QY 362 ArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsn 381
 DB 1795 -----CTAAGACGACGACAAACAGCATATAATAGTAGTAGAAAT 1836

RESULT 17
 US-09-134-001C-322
 ; Sequence 322, Application US/09134001C
 ; Patent No. 6380370


```

; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 322
; LENGTH: 30549
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-322

Alignment Scores:
Pred. No.: 0.0062 Length: 30549
Score: 133.00 Matches: 73
Percent Similarity: 37.96% Conservative: 50
Best Local Similarity: 22.53% Mismatches: 115
Query Match: 5.85% Indels: 87
DB: 4 Gaps: 14

US-10-057-531A-2 (1-431) x US-09-134-001C-322 (1-30549)
QY 12 valproargGlySerGlyMetLysGluThrAlaAlaLysPheGluArgGlnHisMet 31
||||| ||||| : : : : :
Db 12859 GTACCTAGTACAGATGCTATGACTCAGCAATCTGTTGCAAAATATATATCAAAACTACAA 12918
QY 32 -----AspSerProAspLeuGly 37
: : : : :
Db 12919 ATAGCTAAAAACGAAATTACACAATTAATTAACGTTTACCGAACAAATCCAGATGTTAAT 12978
QY 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57
: : : : :
Db 12979 GCAATCAAAACGAATAAGCAGAACGAA----- 13008
QY 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
: : : : :
Db 13009 -----CGAATCAGTAACGATTTAAACACAGCTAAGATAACTTACAAAGTT--- 13053
QY 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
: : : : :
Db 13054 GATACTCAACCTTAGAAAAAATAAAGACAACTTCAAGATGAAATGTATCAAGGTACT 13113
QY 98 PheThrPheAsnLeu-----AsnLeuAsnAspIleLeu----- 108
: : : : :
Db 13114 AACACAGATGGAATGACTCAAGATTCAGTGGATAATTACAATGATAGTCTTAAGTGCAGCA 13173
QY 109 -----AsnSerArgLeuLysLysArgLysTyrPheLeuAspVal 121
: : : : :
Db 13174 ATTATAGAAAAGCAAGTAATAATAATTAATTAACGTAATCCGACAGTAGAACAAAGTT 13233
QY 122 LeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAsp 141
: : : : :
Db 13234 AAAGAGAGC-----GTTGCTAATGCACAAACAGTCATACAGAT 13272
QY 142 SerPheLysLeuLeuAsnSer-----GluGlnLysAsn 152
: : : : :
Db 13273 TTACAAAATGCTCGAAGCTTCACTTGTTCACAGAAAAACTCAACTTCAAGAGCTAAAAAT 13332
QY 153 ThrLeuLysSerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAla 172
: : : : :
Db 13333 AGATTAGAAAACAGT-----ATTAAACCAACAAACAGATACACTGAC---GGCATGACT 13380
QY 173 GlnGluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSer 192
: : : : :
Db 13381 CAAGATTCGCTTACATATTATATGATAAATGACAAAGCTAGACAAACCTTGAAAAA 13440
QY 193 IleLysLysValIleLysGluLysGluLysPheProSerProThrPro 212

```

```

Db 13441 ATATCTAAAGTTTA-----GGTGGCTCAACCTACTGTAGCT 13476
QY 213 ProSerProAlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThr 232
Db 13477 GAAATTAGACAAAATACAGATGAAGCAAAATGCACATAAACAAGCATTA----- 13524
QY 233 AsnIleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeu-----Ile 250
Db 13525 -----GACACTGCACGCTTCTCAACTTACATTAATAGAGAGCCATATATCAATCATATT 13578
QY 251 AsnLeuLysAlaLysIleAsnAspCysAsnValGluLysAspGluAlaHisValLysIle 270
Db 13579 AATAATGAAAGTCATTTAAAT-----AACGCGCAAAAGATAATTTTAAAGCTCAAGTT 13632
QY 271 ThrLysLeu-----SerAspLeuLysAlaIleAspAspLysIleAspLeuPheLys 287
Db 13633 AACTCAGCACCTAATCATATACTTTAGAAACGATTAAATAAGGCTGATAC-TTTAAA 13691
QY 288 AsnProTyrAsp 291
Db 13692 TCAATCTATGAC 13703

RESULT 18
US-07-792-865D-1
; Sequence 1, Application US/07792865D
; Patent No. 5646247
; GENERAL INFORMATION:
; APPLICANT: John W. Barnwell, Mary W. Galinski,
; APPLICANT: Samuel P. Wertheimer
; TITLE OF INVENTION: MEROZOITE ANTIGENS LOCALIZED AT
; TITLE OF INVENTION: THE APICAL END OF THE PARASITE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby P.C.
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM or IBM-compatible
; OPERATING SYSTEM: PC/MS-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07792,865D
; FILING DATE: 19911004
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US90/01849
; FILING DATE: April 3, 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson, Arthur
; REGISTRATION NUMBER: 34,354
; REFERENCE/DOCKET NUMBER: 5986/14692-US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700
; TELEFAX: (212) 753-6237
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3763 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; DESCRIPTION:
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: P.vivax
; STRAIN:

```


QY 427 IlePheCys 429
Db 3705 ATTAATGC 3713

RESULT 19
US-08-468-036-33
; Sequence 33, Application US/08468036
; Patent No. 5728806
; GENERAL INFORMATION:
; APPLICANT: DeMaggio, Anthony J.
; APPLICANT: Hoeksra, Merl F.
; TITLE OF INVENTION: Materials and Methods Relating to Proteins that
; INTERACT WITH CASEIN KINASE I
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,036
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/184,605
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5728806and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/31784
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3883 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-468-036-33

Alignment Scores:
Pred. No.: 0.000916 Length: 3883
Score: 128.00 Matches: 89
Percent Similarity: 39.07% Conservative: 54
Best Local Similarity: 24.32% Mismatches: 109
Query Match: 5.63% Indels: 114
DB: 1 Gaps: 21

US-10-057-531A-2 (1-431) x US-08-468-036-33 (1-3883)

QY 86 TyrArgSerLeuLysGlnIleGluLysAsnIlePheThrPheAsnLeuAsn 105
Db 2011 TATGCTTCGAAGGCTAAACATTAAGAACACCGCAAGCTGGTTCATTATAATGAAG 2070

QY 106 AspIleLeuAsnSerArgLeuLysArgLysTyrPheLeuAspValLeuGluSerAsp 125
Db 2071 GATATTG-----GTTAAATAATACTATGGAATTAGCAAGATTAAATCCGAT 2121

QY 126 LeuMetGlnPheLysHisIleSerSerAsnGlu-----TyrIleIleGluAspSerPhe 143
Db 2122 TTACTCTCTACAAG-----TCCAAAGAGGAATATATATGAGCCAGATCACTAC 2172

QY 144 LysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLysGlu 163

Db 2173 AAAAATTTGAACAGTGAT-----TTAGAAAGTTATAAAATGAAGTTCAA 2217
QY 164 SerValGluAsnAspIleLysPheAlaGlnGluClyIleSerTyrTyrGluLysValLeu 183
Db 2218 GAATGTAAGAGAGAAATGAA-----AGTTTGACATCGAAAAATGCA 2259
QY 184 AlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluLysGluLys 203
Db 2260 TTGCTAGTAAAGATAAATGAAGTCA--AAAGAACTATTCAATCTCAAAATGCGAA 2316
QY 204 PheProSerSerProProThrThrProProSerProAlaLysThrSerGluGlnLysLys 223
Db 2317 ATAGAATCATTTGAAAACTACCATAGATCATTTAAGGCGACAACTAGATAAACAACATA 2376
QY 224 GluSerLysPheLeuProPheLeuThrAsnIleGlu---ThrLeuTyrAsnAsnLeuVal 242
Db 2377 -----ACTGAAATTGAAATATCCCGATTATTAATAACAACTA 2412
QY 243 AsnLysIleAspAspTyr-----LeuIleAsnLeuLysAlaLysIleAsnAsp 258
Db 2413 CAGAAGTTGACTGAGGTAATGCAAAATGCCCTACATGATTACAAAAAAGAGAACTTGAC 2472
QY 259 CysAsnValGluLysAspGluAlaHisValLysIleThrLys---LeuSerAspLeuLys 277
Db 2473 CTTAAT---CAAAAGTTTGAATGTCAT-----ATTACTAAAGAAATTAATAAATGAAA 2523
QY 278 Ala----- 278
Db 2524 TCTACACTGTTTACAAATTAACACTATGCAACAGAGAAAGTATTCTTCAAGAGACTAAT 2583
QY 279 IleAspAspLysIleAspLeuPheLysAsnPro----- 289
Db 2584 ATCCAACCAATCTTGATATGATCAAAATGAAGTACTGACTCTTATGAGAACCATGCAA 2643
QY 290 -----TyrAspPheGluAlaIleLysLysLeuIleAsnAspAspThr 303
Db 2644 GAAAAAGCTGAACATAATGTACAAA---GACTGTGTGAAGAAATTTTAAACGAATCTCCT 2700
QY 304 LysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThr 323
Db 2701 AAA-----TTCTTCAATGTT 2715
QY 324 IleIleSerLys-----LeuIleGluGlyLysPheGln----- 334
Db 2716 GTTATTGAGAAATCGACATAATAAGAGTAGATTTCCTCAAAATTTTATAAAATATAGCC 2775
QY 335 ---AspMetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSer 353
Db 2776 GAGAATCTTCTGATATTAGCGAAGAAATAACAACTGAACAGTAGTACTTTAAAAAACCAT 2835
QY 354 GlyCysPhe-----ArgHisLeuAspGluArgGlu 363
Db 2836 -----TTTTTCAGATAATACCATCAAGAATTACTGAATCGTCATGTGGATTCTACTAT 2889
QY 364 GluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsn 383
Db 2890 GAA-----AATATTGAGAAGAAACAACGAGTTTGTGTGAGAACTTTAAA 2934
QY 384 ProThrCysAsnGluAsn 389
Db 2935 AAGGTCCTAAATGACCAC 2952

RESULT 20
US-08-376-843-33
; Sequence 33, Application US/08376843
; Patent No. 5846764
; GENERAL INFORMATION:
; APPLICANT: DeMaggio, Anthony J.
; APPLICANT: Hoeksra, Merl F.
; TITLE OF INVENTION: Materials and Methods Relating to Proteins
; INTERACT WITH CASEIN KINASE I
; NUMBER OF SEQUENCES: 53

[illegible]

US-10-057-531A-2 (1-431) x US-08-376-843-33 (1-3883)

```

RESULT 21
US-09-541-782-3
; Sequence 3, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541,782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 3884
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-541-782-3

Alignment Scores:
Pred. No.: 0.000916
Score: 128.00
Percent Similarity: 39.07%
Match:
Conser

```


Qy	243	AsnLysIleAspAspTyr-----	LeuIleAsnLeuLysAlaLysIleAsnAsp	258
		: : : :	: : : : : :	
Db	2414	CAGAAGTCTGACTGAGTAATGCAAAATGGCCCTACATGATTACAAAAGAGAGACTTGAC	2473	
Qy	259	CysAsnValGluLysAspGluAlaHisValLysIleThrLys--	LeuSerAspLeuLys	277
		: : : : : :		
Db	2474	CTTAAT---CAAAAGTTTGAATGCAAT-----	ATTACTAAAGAAATTAATAAAATTCGAAA	2524
Qy	278	Ala-----		278
		: : :		
Db	2525	TCTACACTGTTTTACAATTAACACTATGCAACAGGAAAGTATTCCTCAAGAGACTAAT	2584	
Qy	279	IleAspAspLysIleAspLeuPheLysAsnPro-----		289
		: : :		
Db	2585	ATCCAACCAAAATCTTGATATGATCAAAAATGAAGTACTGCCTCTATTATGAGAACCATCGAA	2644	
Qy	290	-----TyrAspPheGluAlaIleLysLysLeuIleAsnAspAspThr	303	
		: : : : : : : : : : : : : : : : : :		
Db	2645	GAANAAGCTGAAGCTAATGTACAAA--	GACTGTGTGAGAGAAATTTTAAACGAACTCTCCT	2701
Qy	304	LysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThr	323	
		: : : : : : : : : : : : : : : : : :		
Db	2702	AAA-----	TTCTTCAATGTT	2716
Qy	324	IleIleSerLys-----	LeuIleGluGlyLysPheGln-----	334
		: : : : : : : : : : : : : : : :		
Db	2717	GTTATTGAGAAATCGACATATAAGAGTAGATTTCCTCAAAAATTTTATAAAAATATAGCC	2776	
Qy	335	---AspMetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSer	353	
		: : : : : : : : : : : : : : : : : :		
Db	2777	GAGAACTTTCTGATATTAGCGAGCAAGAAATACACACATGAACAGTACTTAAAAAACCAT	2836	
Qy	354	GlyCysPhe-----	ArgHisLeuAspGluArgGlu	363
		: : : : : : : : : : : : : : :		
Db	2837	-----TTTTTCAAGAATAACCATCAAGAANTTACTGAAATCTCATGTGGATTCTACTTAT	2890	
Qy	364	GluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsn	383	
		: : : : : : : : : : : : : : : : : :		
Db	2891	GAA-----	AATATTGAGAGAGACAAACGAGTGTGTGTGAGAACTTTAAA	2935
Qy	384	ProThrCysAsnGluAsn	389	
		: : : : : : : : : : : : : : :		
Db	2936	AAGGTCTCTAAATGACCAC	2953	

Percent Similarity:	38.50%	Conservative:	68
Best Local Similarity:	21.50%	Mismatches:	133
Query Match:	5.60%	Indels:	113
DB:	4	Gaps:	19
US-10-057-531A-2 (1-431) x US-09-157-257-3 (1-2836)			
Qy	49	GlySerIleGluGlyArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSer	68
Db	415	GGACTTATAGACGGTAAATGTTAGACGGAGGCCAAAGACAGTACGCACCTAAT	465
Qy	69	GlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSer	88
Db	466	-----GAAAAATTCAGAGGAGCTTTTTCGAAGC	492
Qy	89	LeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeu---AsnLeuAsnAspIle	107
Db	493	AATGGTAAAGCACATACTGAAGACATCTTACTGATPAACACCGGTAACTTTAAAGCGCCT	552
Qy	108	LeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValIleGluSerAspLeuMet	127
Db	553	ATAGAAAGTCGACCTAAGGTAG-----CTGAAGATCTCTTATTGATGAA-----	600
Qy	128	GlnPheLysHisIleSerSerAsnGlu-----TyrIleIleGlu	140
Db	601	AAATTTCAAAAATATTTCGAGGATGAACGAAAGCTGGTCTGTTAAAGAAATACTTACA	660
Qy	141	AspSer-----PheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyr	158
Db	661	GACAGCAACGCTAAGGAAATACTCAACAATGAAGTAGACAAAAGAGGTACTAAATCCGAT	720
Qy	159	LysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyr	178
Db	721	AAATTC---AAGGAGGCAATACTGGCGATGGTAAAGCGCCTAAAGACAGATCTTACT	777
Qy	179	TyrGluLysVal-----	182
Db	778	TGCTATAAATTTAAAGAGGCTGTAAACAGCAATGGTAAAGACATACTAAAGGATATACTT	837
Qy	183	-----LeuAlaLysTyrLysAspAspLeuGluSer-----IleLys	194
Db	838	ACAGATAGCACTGGTAAATTTAAAGAACTTATAGAAGGTACTAGTAAAGACATACTAAA	897
Qy	195	LysValIleLysGluGluLysGluLysPheProSerSerProProThrThrProProSer	214
Db	898	GAGATACTACAGATAATACCGTAACTTTAAAGGCCTTATAGAAAGCACT-----GGC	951
Qy	215	ProAlaLysThrAspGlnLysLysLysGluSerLysPheLeuProPheLeuThrAsnIle	234
Db	952	AAGGAGAAAGCTAAAGAACTCTTATCGATGGGAAGCTTTAAAGGACCTGTTTACTGATGCA	1011
Qy	235	-----GluThrLeuTyrAsnAsnLeuValAsnLysIleAspasp	247
Db	1012	ACAAAGCCGGTTATGTAAAGAAATTACTCAAGACGATACAGCTAAGGAAGTA-----	1055
Qy	248	TyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGluLysAspGluAlaHis	267
Db	1066	--CTTACAGATCAACACAGCAAG-----GAGGTCTTAAAGATAGTACACCT	1110
Qy	268	ValLysIleThrLysLeuSerAspLeuLysAlaIleAspLysIleAspLeuPheLys	287
Db	1111	AAAGACATATTAAGGACACAAACGCGTAAAGAAAGTGCTTACAGATGGT	1155
Qy	288	AsnProTyrAspPheGluAlaIle-----LysLysLeuIleAsnAspasp	302
Db	1153	AACAGCACAGCTAAGAAATATCTACAAACCAACCGCTAAAGAAAGTGCTTACAGATGGT	1212
Qy	303	ThrLysLysAspMetLeuGlyLysLeuLeu-----	312
Db	1213	ACATCCAAAGAAGTACTAAAGAGATACTTACTTGTGATAAATTTAAAGAGCGCATACA	1272
Qy	313	-----SerThrGly-----	315

Db	1273	GGAGATGGTAAAGACATACTAAAGCGTATACCTTACAGATACGACTGGTAAAGTAAAGAA	1332
Qy	316	LeuValGlnAsnPheProAsnThrIleIleSerLysLeuLe-----GluGlyLys	332
Db	1333	CTTATAGAAGTACTGGTAAACACATACTGAAGACATCTCTTACAGATACGACTGGTAA	1392
Qy	333	PheGlnAspMetLeuAsnLleSerGlnHisGlnCysValLysLysGlnCysProGluAsn	352
Db	1393	TTTAAAGAACTTATAGACGTCTGCTGTAAAGAACACAGCTAAAGAGATTCCTTACAGATAAC	1452
Qy	353	SerGlyCysPheArgHisLeuAspGlu-----ArgGluGluCysLysCysLeuLeu	369
Db	1453	ACCGGTAACTTCAAGGGCTCTTGAAGACGCCCGGGAAGGATGAAGCAAAAGCACTACT	1512

```

RESULT 24
US-09-134-001C-1541
; Sequence 1541, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GFC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1341
; LENGTH: 2637
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1541

```

Alignment Scores:	
Pred. No.:	0.000737
Score:	126.50
Percent Similarity:	37.75%
Best Local Similarity:	22.82%
Query Match:	5.56%
DB:	4
Length:	2637
Matches:	81
Conservative:	53
Mismatches:	116
Indels:	106
Gaps:	16

US-10-057-531A-2 (1-431) x US-09-134-001C-1541 (1-2637)

Qy	73	GlutTyraSpValileTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGln	92
Db	1391	AACTACAGGGTAAATATCTAGAGAAG---AACAATACATACAGAATACTAAACAAATT-	1446
Qy	93	IleGluLysASN-----IlePheThrPheAsnLeuAsnLeuAsn	105
Db	1447	ATTGAAGACAAATGATATTTCTATAACAAATATCATTTTACCTTTTACGCAAAATTTATAAT	1506
Qy	106	AspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeuGluSerAsp	125
Db	1506	-----	1506
Qy	126	LeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSerPheLysLeu	145
Db	1507	---CAACAAATTAACATCTTCAG-----ATTTTGTGATAGTATTAAAAAT	1548
Qy	146	LeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyr-----Ile	161
Db	1549	ATT--ACAGACAATCATATAAAAAATTTAAAGAAATAAAATCTATCTTTTACTGACCTA	1605
Qy	162	LysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyr-TyrGluLys	181
Db	1606	AAAGATACTACACAAACCAATA-----GAAAA	1635
Qy	182	ValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValIleLys-----	198
Db	1636	ATTATATATCTTTGGAAGACAAAGAAAAAATATTGAGAGGAAATCAATAGACGTATT	1695

Db 1295 CTGAACAAATGGAATAATGAA-----TACAAACAAAAATGAGAAATTTGAAATTT 1345
QY 110 SerArgLeuLys-----LysArgLysTyrPheLeuAspValLeuGlu----- 123
Db 1346 ATCAAGATTAAACAGTTTGAAGAGCGTCGCTTTAGATAAAATAGAAGAGGTA 1405
QY 124 -----SerAspLeuMetGln 128
Db 1406 AGAAATAAATACCATGAACCTTCCACTTTACAGGAATGTTGAACGATGTTGAACAA 1465
QY 129 PheLysHisIleSerSerAsnGluTyrIleGlu---AspSerPheLysLeuLeuAsn 147
Db 1466 AACCATATGCTTGAAGAGAGAAATGCGCTTACAGATACCAATCGCGTGAAGAAAGGAT 1525
QY 148 SerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLysGluSerValGluAsn 167
Db 1526 ATAGAGCTGAATAATAACATATGCAAGAAATCGAAAGCATAAAAGAGAAATCGAAAT 1585
QY 168 AspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLys 187
Db 1586 ACATTAATA-----CCTGAGTTGGCAGAAAAAGAAAGCTCTTAACAGAAAAAGCGT 1636
QY 188 AspAspLeuGluSerIleLysValIleLysGluGluLysGluLysPheProSerSer 207
Db 1637 AACCGGTATGAGCTATCAAGTTAAAGTTAAAGAAAGGAGAGAA----- 1684
QY 208 ProProThrThrProProSerProAlaLysThrAspGluGlnLysLysGluSerLysPhe 227
Db 1685 -----ACTACAGGCTGAGAGATGAGGTGGCATTAAACACAGAA-- 1723
QY 228 LeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAsp 247
Db 1724 -----ACTAATTTAGAACTTTG-----GAAAGATCAAGAACTTGAGGAA 1765
QY 248 TyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn-----ValGluLysAspGlu 265
Db 1766 TATATAAGACACTGAAGCTGGTATGAGGAGTTGAATGAAATTCGATTAAGAGGAA 1825
QY 266 AlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAspAspLysIleAspLeu 285
Db 1826 ACG---GTTAGCGCACATTGCTAATAGTACAGAGTTAAGAGGAAATATACGAGTT 1882
QY 286 Phe-----LysAsnProTyrAspPheGluAlaIleLysLysLysIleAsnAspAspThr 303
Db 1883 TATTGTAGGATTCGTCCA-----GCTCTAAAAAATTTGGAAAAATTCGTACT 1930
QY 304 LysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThr 323
Db 1931 AGCCTTATTATGTTAATGATTTGATGACAATAGTGGTGTTCATATCTATGGAGTGACG 1990
QY 324 IleIle-----SerLysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSer 340
Db 1991 AAAATACAAAACACAGCGCAAGTGCATGATTCAAATTTTCATAAAATATTTGAT----- 2044
QY 341 GlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAsp 360
Db 2045 -----CAACAGGATACAAATGTGCGATGTTTAAAGAAAGTTGGT 2083
QY 361 GluArgGluGluCysLys-----CysLeuLeuAsnTyrLysGlnGlu 374
Db 2084 CAGTTAGTGCAGAAATTCATTAGATGATGATATGTTTGTATCTTCGCATACGCNCAACA 2143
QY 375 Gly 375
Db 2144 GGA 2146

RESULT 27

US-08-376-843-38
; Sequence 38, Application US/08376843
; Patent No. 5846764
; GENERAL INFORMATION:
; APPLICANT: DeHagglo, Anthony J.
; APPLICANT: Hoekstra, Meri F.

; TITLE OF INVENTION: Materials and Methods Relating to Proteins
; TITLE OF INVENTION: that Interact with Casein Kinase I
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/376,843
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/184,605
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5846764and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/31784
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3466 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-376-843-38

Alignment Scores:

Score: 0.00127 Length: 3466
Percent Similarity: 126.00 Matches: 73
Best Local Similarity: 41.94% Conservatives: 70
Query Match: 21.41% Mismatches: 126
Indels: 72
DB: 5.54% Gaps: 15

US-10-057-531A-2 (1-431) x US-08-376-843-38 (1-3466)

QY 70 PheGluAsnGluTyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeu 89
Db 1235 TTCATGAATCAAGGACATGAATACATTTAAAGCAGCTAGCGGCATCTAATAAAAAAGAG 1294
QY 90 LysLysGlnIleGluLysAsnIlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsn 109
Db 1295 CTGAACAAATGGAATAATGAA-----TACAAAAACAAAAATTGAGAAATTTGAAATTT 1345
QY 110 SerArgLeuLys-----LysArgLysTyrPheLeuAspValLeuGlu----- 123
Db 1346 ATGAGATTAAACAGTTTGAAGAAATGAAGAGCGTCGCTTTAGATAAAATAGAAGAGGTA 1405
QY 124 -----SerAspLeuMetGln 128
Db 1406 AGAAATAAATCACCATGAACCTTCCACTTTACAGGAATGTTGAACGATGTTGAACAA 1465
QY 129 PheLysHisIleSerSerAsnGluTyrIleIleGlu---AspSerPheLysLeuLeuAsn 147
Db 1466 AACCATATGCTTGAAGAGAGAAATGCGCTTACAGATACCAATCGCGTGAAGAAAGGAT 1525
QY 148 SerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLysGluSerValGluAsn 167
Db 1526 ATAGAGCTGAATAATAACATATGCAAGAAATCGAAAGCATAAAAGAGAAATCGAAAT 1585


```
Db 2490 GATCAATTCGGTTA-----AATCCATGTACCATAGAATAATA-----TTATTATCT 2537
Qy 301 AspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuValGlnAsnPhe 320
Db 2538 TCAGCAACTATAAA-----TCAAACATATTAGCGTTACGGCGCAGTAAATACTGG 2588
Qy 321 -----ProAsnThrIleIleSerLysLeuIleGluGlyLysPhe---Gln 334
Db 2589 AAATGTTATTTCGCAAAATGTATCAATGTATATAAAAAATAAAGGTTGTATCGTAATG 2648
Qy 335 AspMetLeuAsnIleSerGlnHisGlnCysValLys-----LysGlnCysProGluAsn 352
Db 2649 GATATGGTTGATATATATCTACTAACATCTTAAATACCATAAACAATATATGATAAA 2708
Qy 353 SerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCys----- 367
Db 2709 ATGAGTACGTTGCAATATATAACGAGATATTAATCATGTAAATGCTCGATATGTTCCGAC 2768
Qy 368 ---LeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsnProThrCys 386
Db 2769 TCTATAACACATCATATATGAAACAACATCATGTATA---AATTATAAATCTACCGAT 2825
Qy 387 AsnGlu 388
Db 2826 AATGAT 2831

RESULT 31
US-08-375-992A-113
; Sequence 113, Application US/08375992A
; Patent No. 6328975
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 220
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,992A
; FILING DATE: Herewith
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3628 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Swinepox virus
; STRAIN: Kasza
; INDIVIDUAL ISOLATE: S-SPV-001
; IMMEDIATE SOURCE:
; CLONE: 515-85.1
```

```
; POSITION IN GENOME:
; MAP POSITION: ~23.2
; UNITS: %G
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 57..1226
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1362..3395
; US-08-375-992A-113

Alignment Scores:
Pred. No.: 0.00136 Length: 3628
Score: 126.00 Matches: 78
Percent Similarity: 41.81% Conservative: 65
Best Local Similarity: 22.81% Mismatches: 135
Query Match: 5.54% Indels: 64
DB: 4 Gaps: 19

US-10-057-531A-2 (1-431) x US-08-375-992A-113 (1-3628)
Qy 75 AspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGlu 94
Db 1914 GATGCATTCACATACCCCCACACATATATCATTCAGTTTCACCT---AGAGATATAAAC 1970
Qy 95 LysAsnIlePheThrPhe-----AsnLeuAsnLeuAsnAspIleLeuAsnSer 110
Db 1971 AGGATATTTCGAATTCGTTAAAAAATATCCGAATAATAATATTATTCATATATATCCGAT 2030
Qy 111 ArgLeuLysLysArgLysTyrPheLeuAspValLeuGlu-----SerAspLeuMet 127
Db 2031 AGCATAAATCAAAATAGTTTCATTCATTCACATATGATGATATCAATATCAATATGTTT 2090
Qy 128 GlnPheLysHisIleSerSerAsnGluTyrIle-----IleGluAspSerPheLysLeu 145
Db 2091 CTGCTATATATCCCTAGTGTAAACGATTTTATATCATCCGATGTTGATAAAGATCGACTT 2150
Qy 146 LeuAsnSerGluGln---LysAsnThrLeuLeuLysSerTyrLysTyr----- 160
Db 2151 ATTAATATGTATGGGATTAAAGTGTGTTGCTATGTTTCGTACGATATAAATCATGATCGAT 2210
Qy 161 IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyLysSerTyrTyrGlu 180
Db 2211 TTAGAGTCATTAGATCAGTACAGATTATATTTATAGAAAAAATATATATATATATAGAC 2270
Qy 181 LysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluGlu 200
Db 2271 -----GTTAAATGTAGAGATTTTCCGAATATGATTTAGAGATAGGTTAAAGAGAA 2321
Qy 201 LysGluLysPheProSerSerProThrThrProProSerProAlaLysThrAspGlu 220
Db 2322 AAGAATAGAAATATAACTACG-----AAATGTGAAGAT 2354
Qy 221 GlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsn 240
Db 2355 ATTATAAGATATATAAAATTTATTCAGT-----AAAAATAGAAATAAAGCAT 2399
Qy 241 LeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn 260
Db 2400 GAAAAATAAAGTGGAGGAGGTGTTGATACATATTGAT-----AAT 2441
Qy 261 ValGluLysAspGluAlaHisValLysIleThrLysSerAspLeuLysAlaIleAsp 280
Db 2442 GTATCTAAAAATAATAAA-----TTATCACTGTCGATATATCATCTTTAATG 2489
Qy 281 AspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsn 300
Db 2490 GATCAATTTTCGTTTA-----AATCCATGTACCAATAAGAAATAATA-----TTATTATCT 2537
Qy 301 AspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuValGlnAsnPhe 320
Db 2538 TCAGCAACTATAAA-----TCAAAACTATTACGTTACGGCGCAGTAAATACTGG 2588
```

```

QY 321 -----ProAsnThrIleIleSerLysLeuIleGluGlyLysPhe---Gln 334
Db 2589 AAATGTTTATTCATTGACAAATGATCAATGATATAAAAGGGTGTTCGTAATG 2648
QY 335-AspMetLeuAsnIleSerGlnHisGlnCysValLys-----LysGlnCysProGluAsn 352
Db 2649 GATATGGTTGATTATATCTACTAACATTTCTTAATACCATAAACAATTATATGATAA 2708
QY 353 SerGlyCysPheArgHisLeuAspGluArgGluGlyCysLysCys----- 367
Db 2709 ATGAGTACGTTTGAATATAAAGCAGATATAAATCATGCTAAATGCTCGATATGTTCCGAC 2768
QY 368 ---LeuLeuAsnThrLysGlnGluCysLysCysValGluAsnProAsnProThrCys 386
Db 2769 TCTATAACATCATATATATGAAACAACATCATGATATA---AATTATAATCTACCGAT 2825
QY 387 AsnGlu 388
Db 2826 AATGAT 2831
RESULT 32
US-08-480-640A-189
; Sequence 189, Application US/08480640A
; Patent No. 6033904
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,640A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 189:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3942 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Swinepox virus
; STRAIN: Kasza
; INDIVIDUAL ISOLATE: S-SPV-001
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..369
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 370..597
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 598..1539

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1675..3708
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (3748..3942)
US-08-480-640A-189
Alignment Scores:
Pred. No.: 0.00155 Length: 3942
Score: 126.00 Matches: 78
Percent Similarity: 41.81% Conservative: 65
Best Local Similarity: 22.81% Mismatches: 135
Query Match: 5.54% Indels: 64
DB: 3 Gaps: 19
US-10-057-531A-2 (1-431) x US-08-480-640A-189 (1-3942)
QY 75 AspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIleGlu 94
Db 2227 GATGCAATTCACATACCCCCACACATATATCATTCATTCCTCCTT---AGAGATATAAAC 2283
QY 95 LysAsnIlePheThrPhe-----AsnLeuAsnLeuAsnAspIleLeuAsnSer 110
Db 2284 AGGATTATTGAATTGCTTAAAAAATATCCGAATAATAATATATTGATTATATATCCGAT 2343
QY 111 ArgLeuLysLysArgLysTyrPheLeuAspValLeuGlu-----SerAspLeuMet 127
Db 2344 AGCATAAATCAATAGTTTCATTCATTCACATCTTCATATGATATAATCAATATGTTT 2403
QY 128 GlnPheLysHisIleSerSerAsnGluTyrIle-----IleGluAspSerPheLysLeu 145
Db 2404 CCTGCTATAATCCCTAGTGTAAAGGATTTTATATCTACCGTAGTTGATAAAGATCGACTT 2463
QY 146 LeuAsnSerGluGln---LysAsnThrLeuLeuLysSerTyrLysTyr----- 160
Db 2464 ATTAATATGATGGGATTAAGTGTGCTATGTTTCGTACGATATAAAGATCATCGAT 2523
QY 161 IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGlu 180
Db 2524 TTAGAGTCATTAGATGACTCAGATTACATATTTATAGAAAAAATATATCTATATACGAC 2583
QY 181 LysValLeuAlaLysTyrLysAspAspLeuSerIleLysLysValIleLysGluGlu 200
Db 2584 -----GTTAAATGTAGAGATTTTCCGAATATGATTAGAGATAAGGTTAAAGAGAA 2634
QY 201 LysGluLysPheProSerSerProThrProProSerProAlaLysThrAspGlu 220
Db 2635 AAGAATAGAAATTAATACTACG-----AAATGTGAAGAT 2667
QY 221 GlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsn 240
Db 2668 ATTATAAGATATATAAAATTTATTCAGT-----AAAAATAGATAAAGCAT 2712
QY 241 LeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn 260
Db 2713 GAAATAATAAGTGGAGAGGTTGTCATACATATTGAT-----AAT 2754
QY 261 ValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAsp 280
Db 2755 GTATCTAAAAATAATAAA-----TTATCAGCTGCTGATATATCATCTTTTAATG 2802
QY 281 AspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsn 300
Db 2803 GATCAATTTCCGTTTA-----AATCCATGTACCAATAAGAAATATA-----TTATTATCT 2850
QY 301 AspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuValGlnAsnPhe 320
Db 2851 TCAGCAACTATAAAA-----TCAAAACTATTACGTTACGGCAGTAAAAAACTGG 2901
QY 321 -----ProAsnThrIleIleSerLysLeuIleGluGlyLysPhe---Gln 334
Db 2902 AAATGTTTATTCATTGACAAATGATCAATGATATAAAAGGGTGTTCGTAATG 2961

```

```

QY 335 AspMetLeuAsnIleSerGlnHisGlnCysValLys-----LysGlnCysProGluAsn 352
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2962 GATATGGTTGATTATATATCTACTACATCTTAAATACCAATAAATATATATGATAA 3021
QY 353 SerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCys----- 367
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3022 ATGAGTACGCTTTGAAATATATAACAGAGATATTAATCATGTAAATGCTCGATATGTTCCGAC 3081
QY 368 ---LeuLeuAsnTyrLysGlnGluGluAspLysCysValGluAsnProAsnProThrCys 386
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3082 TCTATACACATCATATATATGAAACAACATCATGTATA---AATTATAAATCTACCGAT 3138
QY 387 AsnGlu 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3139 AATGAT 3144

RESULT 33
US-08-686-968C-189
; Sequence 189, Application US/08686968C
; Patent No. 6221361
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; FILE REFERENCE: 39119-H/JML
; CURRENT APPLICATION NUMBER: US/08/686,968C
; CURRENT FILING DATE: 1996-07-25
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 189
; LENGTH: 3942
; TYPE: DNA
; ORGANISM: Swinepox virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(369)
; NAME/KEY: CDS
; LOCATION: (370)..(597)
; NAME/KEY: CDS
; LOCATION: (598)..(1539)
; NAME/KEY: CDS
; LOCATION: (1675)..(3708)
; LOCATION: (1675)..(3708)
US-08-686-968C-189

Alignment Scores:
Pred. No.: 0.00155 Length: 3942
Score: 126.00 Matches: 78
Percent Similarity: 41.81% Conservative: 65
Best Local Similarity: 22.81% Mismatches: 135
Query Match: 5.54% Indels: 64
DB: 4 Gaps: 19

US-10-057-531A-2 (1-431) x US-08-686-968C-189 (1-3942)

QY 75 AspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGlu 94
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2227 GATGATTCACATACCCACACACATATATCATTCATTCACCT---AGAGATATAAAC 2283
QY 95 LysAsnIlePheThrPhe-----AsnLeuAsnLeuAsnAspIleLeuAsnSer 110
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2284 AGGATTATTGAAATTGCTTAAAAAATATCCGAATAATAATATTATTCATATATATCCGAT 2343
QY 111 ArgLeuLysLysArgLysTyrPheLeuAspValLeuGlu-----SerAspLeuMet 127
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2344 AGCATAAATCAATATGCTTCATTCATTCATTCATCATATGATATATCAATATATGTTT 2403
QY 128 GlnPheLysHisIleSerSerAsnGluTyrIle-----IleGluAspSerPheLysLeu 145
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2404 CCGTCTATAATCCCTGCTAAACGATTTTATATCTACCGTACTTGTATAAAGATCGACTT 2463
QY 146 LeuAsnSerGluGln---LysAsnThrLeuLysSerTyrLysTyr----- 160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 2464 ATTAATATGATGGGATTAAGTGTGCTGCTATGTTTCGTACGATATAAACATGATCGAT 2523
QY 161 IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrGlu 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2524 TTAGAGTCATAGATGACTGACGATTTATATGAAAAAATAATATATCTATATACGAC 2583
QY 181 LysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluGlu 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2584 -----GTTAAATGTAGAGATTTTGGCAATATGATTAGAGATAAGGTTAAAGAGAA 2634
QY 201 LysGluLysPheProSerSerProThrThrProProSerProAlaLysThrAspGlu 220
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2635 AGAATAGATATATACTACG-----AAATGTGAAGAT 2667
QY 221 GlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsn 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2668 ATTATAAGATATATAAATATTTCAGT-----AAAAATAGATAAATACGAT 2712
QY 241 LeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2713 GAAATATAAGTGGAGGAGGTGTTGATACATATTGAT-----AAT 2754
QY 261 ValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAsp 280
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2755 GTATCTAAAAAATAATAA-----TTATCACTGTCTGATATATCATCTTTAATG 2802
QY 281 AspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsn 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2803 GATCAATTTCGTTTA-----AATCCATGTACCATAGAATAATA-----TTATTATCT 2850
QY 301 AspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuValGlnAsnPhe 320
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2851 TCAGCAACTATAAAA-----TCAAAACTATTAGCGTTACGGCAGTAAAAAATGG 2901
QY 321 -----ProAsnThrIleIleSerLysLeuIleGluGlyLysPhe---Gln 334
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2902 AAATGTTATTCATTGACAAATGATCAATGATATAAAAAATAAAGGTTATCGTAATG 2961
QY 335 AspMetLeuAsnIleSerGlnHisGlnCysValLys-----LysGlnCysProGluAsn 352
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2962 GATATGGTTGATTATATATCTACTACATCTTAAATACCATCAACAATATATGATAA 3021
QY 353 SerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCys----- 367
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3022 ATGAGTACGCTTTCATATAAACAGAGATATTAATCATGTAAATGCTCGATATGTTCCGAC 3081
QY 368 ---LeuLeuAsnTyrLysGlnGluGluAspLysCysValGluAsnProAsnProThrCys 386
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3082 TCTATAACACATCATATATATGAAACAACATCATGTATA---AATTATAAATCTACCGAT 3138
QY 387 AsnGlu 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3139 AATGAT 3144

RESULT 34
US-08-488-237A-189
; Sequence 189, Application US/08488237A
; Patent No. 6251403
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```


APPLICATION NUMBER: US/08/375,992A

FILING DATE: Herewith

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: White, John P

REGISTRATION NUMBER: 28,678

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 189:

SEQUENCE CHARACTERISTICS:

LENGTH: 3942 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Swinepox virus

STRAIN: Kasza

INDIVIDUAL ISOLATE: S-SPV-001

FEATURE:

NAME/KEY: CDS

LOCATION: 1..369

FEATURE:

NAME/KEY: CDS

LOCATION: 370..597

FEATURE:

NAME/KEY: CDS

LOCATION: 598..1539

FEATURE:

NAME/KEY: CDS

LOCATION: 1675..3708

FEATURE:

NAME/KEY: CDS

LOCATION: complement (3748..3942)

US-08-375-992A-189

Alignment Scores:

Pred. No.:	Score:	Length:	3942
75	126.00	Matches:	78
2227	41.81%	Conservative:	65
95	22.81%	Mismatches:	135
2284	5.54%	Indels:	64
111	4	Gaps:	19

US-10-057-531A-2 (1-431) x US-08-375-992A-189 (1-3942)

```
QY 75 AspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGlu 94
   ||| :|||: |||
Db 2227 GATGCATTCCACATACCCCCACACATATATCATTCATTACGTTCACTT--AGAGATATAAAC 2283

QY 95 LysAsnIlePheThrPhe-----AsnLeuAsnLeuAsnAspIleLeuAsnSer 110
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2284 AGGATTATTGAATTGCTTTAAAAAATATCCGAATAATAATATTATTATATATATCCGAT 2343

QY 111 ArgLeuLysArgLysTyrPheLeuAspValLeuGlu-----SerAspLeuMet 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2344 AGCATAAATCAATAGTCATTTCATTCATCATTCATTCATATGATATCAAAATATGTTT 2403

QY 128 GlnPheLysHisIleSerSerAsnGluTyrIle-----IleGluAspSerPheLysLeu 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2404 CCTGCTAATACCTAGTGTAAACGATTATATATCTACCGTAGTCTGATAAAGATCGACTT 2463

QY 146 LeuAsnSerGluGln---LysAsnThrLeuLeuLysSerTyrLysTyr----- 160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2464 ATTAATATGTATGGGATTAAAGTGTGTGCTATGTTTTCGTACCATATAAATATGATCGAT 2523

QY 161 IleLysGluSerValGluAsnAspIleLysPheAlaGluGluLysTyrTyrGlu 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2524 TTAGAGTCTATTAGTACGATACGATATTTTATAGAAAAAATATATCTATATACGAC 2583
```

```
QY 181 LysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValIleLysGluGlu 200
   ||| :|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2584 -----CTTAATCTAGAGATTTTCCGAATATGATTAGAGATAGAGTTAAAGAGAA 2634

QY 201 LysGluLysPheProSerSerProThrThrProProSerProAlaLysThrAspGlu 220
   ||| ||| :|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2635 AAGAATAGATATATAACTAGC-----AAATGTTGAAGAT 2667

QY 221 GlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsn 240
   ||| ||| :|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2668 ATTATAGATATATAAATATTTCAGT-----AAAAATAGATAAATACGAT 2712

QY 241 LeuValAsnLysIleAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn 260
   ||| ||| :|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2713 GAAAAATAAAGCTGGAGGAGGTGTCATACATATTGAT-----AAT 2754

QY 261 ValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAsp 280
   ||| ||| :|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2755 GTATCTAAAAAATAATAA-----TTATCACTGCTGATATATCATCTTTAATG 2802

QY 281 AspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsn 300
   ||| ||| :|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2803 GATCAATTTGCTTTA-----AATCCATGATACCATAGAAATATA-----TTATTATCT 2850

QY 301 AspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuValGlnAsnPhe 320
   ||| ||| :|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2851 TCAGCAACTATAAA-----TCAAAACTATTACGTTACGGGCAGTAAAAAACTGG 2901

QY 321 -----ProAsnThrIleIleSerLysLeuIleGluGlyLysPhe---Gln 334
   ||| ||| :|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2902 AAATGTTATTTCATTGACAAATGATCAATGATATAAAAAAATAAAGGCTGTATCGTAATG 2961

QY 335 AspMetLeuAsnIleSerGlnHisGlnCysValLys-----LysGlnCysProGluAsn 352
   ||| ||| :|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2962 GATATGTTGATTATATATCTACTAACATCTTTAAATACCATCAACAAATATATATATAAAA 3021

QY 353 SerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCys----- 367
   ||| ||| :|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3022 ATGAGTACGTTTGAATATAAACGAGATATTAATCATGTAATGCTCGATATGTTCCGAC 3081

QY 368 ---LeuLeuAsnTyrLysGlnGlyAspLysCysValGluAsnProAsnProThrCys 386
   ||| ||| :|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3082 TCTATAACACATCATATATATGAACACATCATGTATA---AATTATAAATCTACCGAT 3138

QY 387 AsnGlu 388
   ||| ||| :|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3139 AATGAT 3144

RESULT 36
US-08-480-640A-221
; Sequence 221, Application US/08480640A
; Patent No. 6033904
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0; Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,640A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
```

QY	75	AspValIleTyrLeuLysProLeualaGlyValTyrArgSerLeuLysLysGlnIleGlu	94
		: : : : :	
Db	4070	GATGCATTCCACATAGCCCCACACATATATCATTTACCTTCACCT---AGAGATAATAAC	4126
QY	95	LysAsnIlePheThrPhe-----AsnLeuAsnLeuAsnAspIleLeuAsnSer	110
		: : :	
Db	4127	AGCATTATTGAATTGGCTTA AAAAATATCCGAATAATAATATTGATTATATATCCGAT	4186
QY	111	ArgLeuLysLysArgLysTyrPheLeuAspValLeuGlu-----SerAspLeuMet	127
		: : : : : :	
Db	4187	AGCATAAATCAANATGTCATTTCATTCACATCTTCATATGTAANTATACAATATGTTT	4246
QY	128	GlnPheLysHisIleSerSerAsnGluTyrIle-----IleGluAspSerPheLysLeu	145
Db	4247	CCTGCTTAATCCCTAGTGTAACAGATTTATATCTACCGTAGTGTGATAAGATCCGACTT	4306
QY	146	LeuAsnSerGluGln---LysAsnThrLeuLeuLysSerTyrLysTyr-----	160
		:: ::	

```
Db 4307 ATTAATATGATGGGATTAAAGTGTGCTGCTATGTTTCGTACGATATAAACATGATCGAT 4366
Qy 161 IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyLeuSerTyrTyrGlu 180
Db 4367 TTAGAGTCATTAGATGACTCGAGTTCATATATTTATGAAAAAATATATCTATATACGAC 4426
Qy 181 LysValLeuAlaLysTyrLysAspAspLeuLysSerIleLysLysValIleLysGluGlu 200
Db 4427 -----GTTAAATGCTAGAGATTTTGGCAATATGATTAGAGATAAGGTTAAAGAGAA 4477
Qy 201 LysGluLysPheProSerProProThrThrProProSerProAlaLysThrAspGlu 220
Db 4478 AAGAATAGATAATTAACAG-----AAATGTGAAGAT 4510
Qy 221 GlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsn 240
Db 4511 ATTATAGATATATAAATTATTCAGT-----AAAAAGATAAATACGAT 4555
Qy 241 LeuValAsnLysLysLeuAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn 260
Db 4556 GAAAAATAAAGGTGGAGGAGGTGTTGATACATATTCAT-----AAT 4597
Qy 261 ValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAsp 280
Db 4598 GTATCTAAAAAATAATAA-----TTATCACTGCTGATATATATCTTTAATG 4645
Qy 281 AspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsn 300
Db 4646 GATCAATTTCGTTTA-----AATCATGTACCATAGAATAATA-----TTATATATCT 4693
Qy 301 AspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuValGlnAsnPhe 320
Db 4694 TCAGCAACTATAAA-----TCAAAACTATTAGCGTTACGGGCGAGTAAAAAATCGG 4744
Qy 321 -----ProAsnThrIleIleSerLysLeuIleGluGlyLysPhe---Gln 334
Db 4745 AATGTTTATTCATGTGCAAAATGATCAATGATATAAAAAAATAAAGGTTATCGTAATG 4804
Qy 335 AspMetLeuAsnIleSerGlnHisGlnCysValLys-----LysGlnCysProGluAsn 352
Db 4805 GATAGTGTGATATATATCTACTAATCAATTCCTTAATATACCATAAACAATATATGATAA 4864
Qy 353 SerGlyCysPheArgHisLeuAspGluArgGluGlyCysLysCys----- 367
Db 4865 ATGAGTACGTTTGAATATAAACAGAGATATTAATCATGTAATGCTCGATATGTTCCGAC 4924
Qy 368 ---LeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsnProThrCys 386
Db 4925 TCATACACATCATATATATGAAACACATCATGTATA---AATTATAAATCTACCGAT 4981
Qy 387 AsnGlu 388
Db 4982 AATGAT 4987
```

RESULT 38

```
US-08-488-237A-221
; Sequence 221, Application US/08488237A
; Patent No. 6251403
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,237A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 221:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: N
; ANTI-SENSE: N
; US-08-488-237A-221
```

```
Alignment Scores:
Pred. No.: 0.00279 Length: 5785
Score: 126.00 Matches: 78
Percent Similarity: 41.81% Conservative: 65
Best Local Similarity: 22.81% Mismatches: 135
Query Match: 5.54% Indels: 64
DB: 4 Gaps: 19
```

US-10-057-531a-2 (1-431) x US-08-488-237A-221 (1-5785)

```
Qy 75 AspValIleThrLysLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIleGlu 94
Db 4070 GATGCAATCCACATACACCCACACATATATCATTCAGTTCACTT---AGAGATATAAAC 4126
Qy 95 LysAsnIlePheThrPhe-----AsnLeuAsnLeuAsnAspIleLeuAsnSer 110
Db 4127 AGGATTTATTCATTCGTTCTTAAAAAATATCCGAATAATAATATTTATGATATCCGAT 4186
Qy 111 ArgLeuLysLysArgLysTyrPheLeuAspValLeuGlu-----SerAspLeuMet 127
Db 4187 AGCATAAAATCAAAATAGTTTCATTCATTCACATATTCATATGATAAATAATATATGTTT 4246
Qy 128 GlnPheLysHisIleSerSerAsnGluTyrIle-----IleGluAspSerPheLysLeu 145
Db 4247 CCTGCTATATCCCTAGTGTAAACGATTTATATCTACCGTAGTTGATAAAGATCGACTT 4306
Qy 146 LeuAsnSerGluGln---LysAsnThrLeuLeuLysSerTyrLysTyr----- 160
Db 4307 ATTAATATGATGGGATTAAAGTGTGCTATGTTTCGTACGATATAAACATCGATCGAT 4366
Qy 161 IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGlu 180
Db 4367 TTAGAGTCATTAGATGACTCGAGTTCATATATTTATGAAAAAATATATCTATATACGAC 4426
Qy 181 LysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluGlu 200
Db 4427 -----GTTAAATGATAGAGATTTTGGCAATATGATTAGAGATAAGGTTAAAGAGAA 4477
Qy 201 LysGluLysPheProSerProProThrThrProProSerProAlaLysThrAspGlu 220
Db 4478 AAGAATAGATAATTAACAG-----AAATGTGAAGAT 4510
Qy 221 GlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsn 240
Db 4511 ATTATAAGATATATAAATTATTCAGT-----AAAAAGATAAATACGAT 4555
Qy 241 LeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn 260
Db 4556 GAAAAATAAAGGTGGAGGAGGAGGTTGTTGATATATTCAT-----AAT 4597
```

Qy 261 ValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAsp 280
Db 4598 GTATCTAAATAATAA-----TTATCACTGCTGATATATCATCTTTAATG 4645
Qy 281 AspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLysIleAsn 300
Db 4646 GATCAATTTGGTTA-----AATCCATGTACCATAAGAAATATA-----TTATATCT 4693
Qy 301 AspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPhe 320
Db 4694 TCAGCAACTATAA-----TCAAACTATTAGGTTACGGGCAGTAAAAACTGG 4744
Qy 321 -----ProAsnThrIleLeuSerLysLysIleGluGlyLysPhe---Gln 334
Db 4745 AAATGTTATTCATTGACAAATGATCAATGATATAAATAAATAAAGGGTCTTATCGTAATG 4804
Qy 335 AspMetLeuAsnIleSerGlnHisGlnCysValLys-----LysGlnCysProGluAsn 352
Db 4805 GATATGGTTGATTATATCTACTAACATCTTAAATACCATAAACAATATATGATAAA 4864
Qy 353 SerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCys----- 367
Db 4865 ATGAGTACGTTTGAATATAAAGCAGATATAAATCATCTGTAATGCTCGATATGTTCCGAC 4924
Qy 368 ---LeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsnProThrCys 386
Db 4925 TCTATAACCATCATATATATGAAACAACATCATGTATA---AATTATAAATCTACCGAT 4981
Qy 387 AsnGlu 388
Db 4982 AATGAT 4987

RESULT 39

US-08-287-959-2
; Sequence 2, Application US/08287959
; Patent No. 5639651
; GENERAL INFORMATION:
; APPLICANT: Weisbach, Lawrence
; APPLICANT: Bernards, Andre
; APPLICANT: Settlement, Jeffrey
; TITLE OF INVENTION: GAP-RELATED GENE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/287,959
; FILING DATE: August 9, 1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul C.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/181001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7573 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

US-08-287-959-2
Alignment Scores:
Pred. No.: 0.00479 Length: 7573
Score: 125.50 Matches: 69
Percent Similarity: 37.97% Conservative: 81
Best Local Similarity: 17.47% Mismatches: 106
Query Match: 5.52% Indels: 139
DB: 1 Gaps: 16
US-10-057-531A-2 (1-431) x US-08-287-959-2 (1-7573)

Qy 75 AspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGlu 94
Db 2889 GAAGTTGTAAGAGATTCACTCCCTGGCAAGGATGCACCAAGCT---CGAAGCGCTATCGA 2945
Qy 95 LysAsnIlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeu----- 112
Db 2946 GATCCCTGCGAGTACTTCCGGGACCATAATAATGACATTATCAAAATCCAGGCTTTTATT 3005
Qy 113 -----LysLysArgLysTyrPheLeuAspValLeuGluSerAsp----- 125
Db 3006 CGGCAACAACAGCTCGGATGACTACAAAGACTCTCATCAATGCTGAGGATCCTCTATG 3065
Qy 126 -----LeuMetGlnPheLysHisIle-----SerSerAsnGluTyrIleIleGluAsp 141
Db 3066 GTTGTGGTCCGAAATTTGTCCACCTGTGGACCAAAAGTACCAGGATTTTCAGGAGGAG 3125
Qy 142 SerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIle 161
Db 3126 CTTGACCTTATGAAGATCGGGAAGAGGTTATCACCTTCATTGCTCT----- 3173
Qy 162 LysGluSerValGluAsnAspIleLysPhe----- 171
Db 3174 AACCAGCAGCTGGAGATGACCTCAATCTCATGGATATCAAAATGGACTGCTAGTGAA 3233
Qy 172 -----AlaGlnGluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLys 187
Db 3234 AATAAGATTAGCTTGCAGGATGTGTTTCCACACAGTAAAAAACTTACCAAAAAAATAG 3293
Qy 188 AspAspLeuGlu-----SerIleLysLysValIle 197
Db 3294 GAACAGTTGCTGATATGATGATGATAATAAACAAGAGGAGGTCTCAAGGCTTTGAGC 3353
Qy 198 LysGluLysGluLysPheProSerSerPro-----ProThrThr 211
Db 3354 AAGGAGAAGAGAGAGAACTTGGAAAGCTTACCAGCACCTGTTTATTATTATTCAAAACCA 3413
Qy 212 ProProSerProAlaLysThrAspGlnLysLysGluSerLysPheLeuProPheLeu 231
Db 3414 CCCACCTATCTGGCCCAAGCTCATTTTCAGATGCCCCAGAACAGTCCACCAAGTTCATG 3473
Qy 232 ThrAsn---IleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIle 250
Db 3474 GACTCTGTAATCTTCACACTCTACAACTACCGGTCCACACGAGGAGGAGTACCTGCTC 3533
Qy 251 -----AsnLeuLysAlaLysIleAsnAspCysAsn 260
Db 3534 CTGGCGCTCTTTAAGACAGCAGCTCCAGAGGAAATCAAGTGAAGGTAGATCAGATTCAA 3593
Qy 261 -----ValGluLysAspGluAlaHisValLys----- 269
Db 3594 GAGATTGTGACAGGAAATCCTACGGTTATTAAATGGTTGAAGTTTCAACCGGGTGCC 3653
Qy 270 -----IleThrLysLeuSerAspLeu 276
Db 3654 CGTGGCCAGAAATGCCCTGAGACAGATCTTGGCCCCAGTCGTGAAGAAATATATGATGAC 3713
Qy 277 LysAlaIleAsp-----AspLysIleAspLeuPheLysAsn----- 288
Db 3714 AAATCTCTCAACATCAAAACTGACCTGTGGATATTTACAAATCTTGGTTAATCAGATG 3773
Qy 289 -----ProTyrAspPheGluAlaIleLysLysLeu 298

Db 3774 GAGTCTCAGACAGGAGGCAAGCAAACTGCCTATGATGTGACCCCTGAGCAGCGCTA 3833
Qy 299 ILeAsnAspThrLysLysAspMet----- 307
Db 3834 GCTCATGAAGAAGTGAAGACACGCTAGACAGCTCCATCAGGAACATGCGGCTGTGACA 3893
Qy 307 ----- 307
Db 3894 GACAAATTTCTCTCAGCCATTGCTGCTGTGGACAAAATCCCTATGGGATGCGCTTC 3953
Qy 308 LeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThr----- 323
Db 3954 ATTGCCAAGCTCTGAAGACTGTTGATGAGAAGTTCCTCGATGCTGCTGAGGATGAG 4013
Qy 324 -----IleIleSerLysLeuIleGluGlyLysPheGlnAsp 335
Db 4014 CTGCTGAAGATTATTGGTGAACCTGCTTTATTATTCATCATCATGAAT 4058
RESULT 40
US-08-235-836C-73
; Sequence 73, Application US/08235836C
; Patent No. 6248562
; GENERAL INFORMATION:
; APPLICANT: Dunn, John J.
; APPLICANT: Luft, Benjamin J.
; TITLE OF INVENTION: No. 6248562el Chimeric Proteins Comprising
; TITLE OF INVENTION: Borrelia Polypeptides and Uses Therefor
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brookhaven National Laboratory
; STREET:
; CITY: Upton
; STATE: NY
; COUNTRY: USA
; ZIP: 11973
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/235.836C
; FILING DATE: 29-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,191
; FILING DATE: 01-11-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Bogosian, Margaret C.
; REGISTRATION NUMBER: 25,324
; REFERENCE/DOCKET NUMBER: BNL93-28A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 282-7338
; TELEFAX: (516) 282-3729
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2107 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2107
US-08-235-836C-73

Alignment Scores:
Pred. No.: 0.000761
Score: 125.00
Percent Similarity: 35.19%
Best Local Similarity: 20.39%
Query Match: 5.49%

Length: 2107
Matches: 84
Conservative: 61
Mismatches: 153
Indels: 114

DB: 4 Gaps: 16
US-10-057-531A-2 (1-431) x US-08-235-836C-73 (1-2107)
Qy 39 AspAspAspLysLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMetAla 58
Db 262 GATCAAGATAAAAGCAAGTGTGATATTTTCTATTGTTAGTAGTAAGTCAGAGCTTGAT 321
Qy 59 IleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyr 78
Db 322 AGTATATTAATCTAAGAAGATTTCTTACAGG-----TAT 357
Qy 79 LeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePhe 98
Db 358 TTAATGAAGTCTTTTATTGAGAGGTCTAGTCGGAATTAATCTCTAAGCATATAC 417
Qy 99 ThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPhe 118
Db 418 ATATATAATGCT----- 429
Qy 119 LeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIle 138
Db 430 -----GTTTATGAGGAGATTTAGATTATTACAA-----GAGTTTAT 468
Qy 139 IleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyr 158
Db 469 ATTGAGGCTTCTTGAAGTCTTTGACTAAAGAAATGCGAGGT---CTTTCTAGGGGTGAC 525
Qy 159 -----LysTyrIleLysGlu 163
Db 526 AGTCAATGGCTGGGAAGACACAAATATTTATTCCTCTTTAAAGAAATATTTATCTGGA 585
Qy 164 SerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLysValLeu 183
Db 586 AATGTTGAGCTGCACATT----- 603
Qy 184 AlaLysTyrLysAspAspLeuGluSerIle-----LysLysValIle----- 197
Db 604 -----GATATTGATAGTTGGTTACAGATAAGGTGCTGCGAGCTCTTTTA 648
Qy 198 -----LysGluGluLysLysPheProSerSerProProThrThrProProSerPro 215
Db 649 AGTGAGATGAATCAGGTGTTAACTTTGCAAGAGATATTACAGACATTCAAGCGCAACT 708
Qy 216 AlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGlu 235
Db 709 CATAAAGCAGATCAAGATAAATTTGATATTGAA-----TTAGATATATTTCAT 756
Qy 236 ThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLysLysAlaLys 255
Db 757 GAAAGTGATTCCCAATATAACAGAAACTATTGAGAATTTAAGGATCAGCTTGAAAAGCT 816
Qy 256 IleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSer--- 274
Db 817 ACAGATGAAGAGCATAAAGAGAGATTGAAGTCAAGGTGATGCTTAAAGAAAGCAAAAG 876
Qy 275 -----AspLeuLysAlaIleAsp-----AspLysIleAspLeuPhe 286
Db 877 GAAGATTAGATAAAAGGCAATTTGATTAAGCTCAACAAATAATGATTTTGTCT 936
Qy 287 LysAsnProTyrAspPhe-----GluAlaIleLysLysLysLeuIleAsn 300
Db 937 GAAGATAATCTAGATATTCAAGAGGATACTGTTAGAGAGAGCTTCAAGAGAAATATTAC 996
Qy 301 AspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPhe 320
Db 997 GAGACTAATAAGGAAAGAAATTTTACCAAG----- 1026
Qy 321 ProAsnThrIleIleSerLysLysIleGluGlyLysPheGln-----AspMetLeuAsn 338
Db 1027 CTGCTGATGTGAAGTTCCTTAAAGTTGATAGCAAGTACAAATAAAGAGAGCTCGAA 1086
Qy 339 IleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHis 358

Db	1087	GATTTCAGGAGCAG	---CTTAAAGAAACTGGTGATGAANAAT			1125
Qy	359	LeuAspGluArgGluGluCysLysCysLeuLeuasnTyrLysGlnGluGlyAspLysCys				378
Db	1126	-----CAGAAAAGAAATTGCAAAAGCAAAATGAAATCAAAAAAAGTGATGAAAAGCTT				1179
Qy	379	ValGluasnProAsnProThrCysAsnGluasnAsnGlyGlyCysAspAlaAspAlaThr				398
Db	1180	TTAAAAGTAAAGATGATAAAGCAAGTAAAGATGCTTAAGCCTTTGGATCTTGATCGAGAA				1239
Qy	399	CysThrGluClnuAspSerGlySerSerArgLysLys				410
Db	1240	TTAAATTCCTAAAGCTTCTAGCAAGAAAAAAGATAA				1275

Search completed: May 19, 2003, 16:17:20
Job time : 139 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 10:16:12 ; Search time 31 Seconds
(without alignments)
409.074 Million cell updates/sec

Title: US-10-057-531A-2
Perfect score: 2275
Sequence: 1 MHHHHHSSGLVPRGSMKE.....TCECTKPSYPLFDGIFCSS 431

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1931	84.9	394	4 US-08-195-705-2	Sequence 2, Appli
2	1921	84.4	394	4 US-08-195-705-4	Sequence 4, Appli
3	1040	45.7	377	4 US-08-195-705-5	Sequence 5, Appli
4	986	43.3	375	4 US-08-195-705-3	Sequence 3, Appli
5	288	12.7	53	1 US-08-290-919-4	Sequence 4, Appli
6	278.5	12.2	824	4 US-09-626-589-3	Sequence 3, Appli
7	277	12.2	48	1 US-08-290-919-12	Sequence 12, Appli
8	275	12.1	467	4 US-09-513-442-2	Sequence 2, Appli
9	274	12.0	48	1 US-08-290-919-2	Sequence 2, Appli
10	274	12.0	167	4 US-09-741-243C-4	Sequence 4, Appli
11	274	12.0	551	4 US-09-741-243C-2	Sequence 2, Appli
12	272	12.0	53	1 US-08-290-919-3	Sequence 3, Appli
13	271	11.9	48	1 US-08-290-919-1	Sequence 1, Appli
14	259.5	11.4	408	2 US-08-683-007A-2	Sequence 2, Appli
15	256.5	11.3	106	1 US-08-290-919-11	Sequence 11, Appli
16	249.5	11.0	209	4 US-09-166-966E-11	Sequence 11, Appli
17	235	10.3	159	4 US-09-166-966E-8	Sequence 8, Appli
18	173.5	7.6	382	4 US-09-277-716-22	Sequence 22, Appli
19	173.5	7.6	382	4 US-09-609-161B-22	Sequence 22, Appli
20	152.5	6.7	288	4 US-09-273-839A-8	Sequence 8, Appli
21	151	6.6	711	3 US-08-946-475-9	Sequence 9, Appli
22	151	6.6	711	3 US-09-340-479-9	Sequence 9, Appli
23	139.5	6.1	1010	4 US-09-134-001C-5178	Sequence 3159, Ap
24	133	5.8	10182	4 US-09-134-001C-3159	Sequence 4, Appli
25	128	5.6	1038	4 US-09-541-782-4	Sequence 4, Appli
26	128	5.6	1038	4 US-09-723-820-4	Sequence 4, Appli
27	127.5	5.6	849	4 US-09-157-257-4	Sequence 4, Appli

28	126.5	5.6	878	4 US-09-134-001C-4378	Sequence 4378, Ap
29	126.5	5.6	1494	3 US-08-755-587-186	Sequence 186, App
30	126	5.5	677	3 US-08-480-640A-115	Sequence 115, App
31	126	5.5	677	3 US-08-480-640A-193	Sequence 193, App
32	126	5.5	677	3 US-08-295-802-115	Sequence 115, App
33	126	5.5	677	4 US-08-686-968C-58	Sequence 58, Appl
34	126	5.5	677	4 US-08-686-968C-193	Sequence 193, App
35	126	5.5	677	4 US-08-488-237A-115	Sequence 115, App
36	126	5.5	677	4 US-08-488-237A-193	Sequence 193, App
37	126	5.5	677	4 US-08-375-992A-115	Sequence 115, App
38	126	5.5	677	4 US-08-375-992A-193	Sequence 193, App
39	125.5	5.5	1657	1 US-08-287-959-1	Sequence 1, Appli
40	125	5.5	700	4 US-08-235-836C-74	Sequence 74, Appl
41	122.5	5.4	2710	1 US-08-480-604A-6	Sequence 6, Appli
42	122.5	5.4	2710	2 US-08-405-496A-6	Sequence 6, Appli
43	122.5	5.4	2710	4 US-08-915-136-6	Sequence 6, Appli
44	122.5	5.4	2710	4 US-08-957-310-6	Sequence 6, Appli
45	119.5	5.3	2285	4 US-09-308-373-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-195-705-2
; Sequence 2, Application US/08195705
; Patent No. 6420523
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra
; APPLICANT: Hui, George
; APPLICANT: Barr, Philip
; APPLICANT: Gibson, Helen
; TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMIDIUM
; TITLE OF INVENTION: FALCIPARUM VACCINE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Davis Hoxie Faithfull Hapgood
; STREET: 45 Rockefeller Pl.
; CITY: New York
; STATE: N.Y.
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,705
; FILING DATE: 14-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jacobs, Seth H
; REGISTRATION NUMBER: 32140
; REFERENCE/DOCKET NUMBER: 11880A3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-757-2200
; TELEFAX: 212-586-1461
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 394 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; STRAIN: falciparum uganda palo alto (FUP)
US-08-195-705-2

```
Query Match      84.4%; Score 1931; DB 4; Length 394;
Best Local Similarity 98.7%; Pred. No. 2.4e-137;
Matches 369; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 58 AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKQIEKNIFTNLFNLNDILNSRLKRRY 117
DB 1 AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKQIEKNIFTNLFNLNDILNSRLKRRY 60
QY 118 FLDVLESOLMQFKHSSNEYIIEEDSFKLLNSQKNTLLKSYKIKESVENDIKFAQEGIS 177
DB 61 FLDVLESOLMQFKHSSNEYIIEEDSFKLLNSQKNTLLKSYKIKESVENDIKFAQEGIS 120
QY 178 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTTPPSAKTDEQKESKFLPFLNIETL 237
DB 121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTTPPSAKTDEQKESKFLPFLNIETL 180
QY 238 YNNLVNKIDDDYLLINLAKINDCNVEKDEAHVKITKLSOLKAIDDKIDLFKNPYDFEAIKK 297
DB 181 YNNLVNKIDDDYLLINLAKINDCNVEKDEAHVKITKLSOLKAIDDKIDLFKNHDFEAIKK 240
QY 298 LINDDTKMDLGLKLLSTGLVQNFNPTIISKLEGFQDMLNISQHCYVKKQCPENSGGCFR 357
DB 241 LINDDTKMDLGLKLLSTGLVQNFNPTIISKLEGFQDMLNISQHCYVKKQCPENSGGCFR 300
QY 358 HLDRECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 417
DB 301 HLDRECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKTEEDSGSNGKKITCECTK 360
QY 418 PDSYPLFDGIFCSS 431
DB 361 PDSYPLFDGIFCSS 374

RESULT 2
US-08-195-705-4
; Sequence 4, Application US/08195705
; Patent No. 6420523
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra
; APPLICANT: Hui, George
; APPLICANT: Barr, Philip
; APPLICANT: Gibson, Helen
; TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMODIUM
; TITLE OF INVENTION: FALCIPARUM VACCINE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Davis Hoxie Faithfull Hapgood
; STREET: 45 Rockefeller Pl.
; CITY: New York
; STATE: N.Y.
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,705
; FILING DATE: 14-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jacobs, Seth H
; REGISTRATION NUMBER: 32140
; REFERENCE/DOCKET NUMBER: 11880A3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-757-2200
; TELEFAX: 212-586-1461
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 394 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; STRAIN: MAD
US-08-195-705-4

Query Match      84.4%; Score 1921; DB 4; Length 394;
Best Local Similarity 98.4%; Pred. No. 1.4e-136;
Matches 368; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 58 AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKQIEKNIFTNLFNLNDILNSRLKRRY 117
DB 1 AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKQIEKNIFTNLFNLNDILNSRLKRRY 60
QY 118 FLDVLESOLMQFKHSSNEYIIEEDSFKLLNSQKNTLLKSYKIKESVENDIKFAQEGIS 177
DB 61 FLDVLESOLMQFKHSSNEYIIEEDSFKLLNSQKNTLLKSYKIKESVENDIKFAQEGIS 120
QY 178 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTTPPSAKTDEQKESKFLPFLNIETL 237
DB 121 YYEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTTPPSAKTDEQKESKFLPFLNIETL 180
QY 238 YNNLVNKIDDDYLLINLAKINDCNVEKDEAHVKITKLSOLKAIDDKIDLFKNPYDFEAIKK 297
DB 181 YNNLVNKIDDDYLLINLAKINDCNVEKDEAHVKITKLSOLKAIDDKIDLFKNHDFEAIKK 240
QY 298 LINDDTKMDLGLKLLSTGLVQNFNPTIISKLEGFQDMLNISQHCYVKKQCPENSGGCFR 357
DB 241 LINDDTKMDLGLKLLSTGLVQNFNPTIISKLEGFQDMLNISQHCYVKKQCPENSGGCFR 300
QY 358 HLDRECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 417
DB 301 HLDRECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSRRKKITCECTK 360
QY 418 PDSYPLFDGIFCSS 431
DB 361 PDSYPLFDGIFCSS 374

RESULT 3
US-08-195-705-5
; Sequence 5, Application US/08195705
; Patent No. 6420523
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra
; APPLICANT: Hui, George
; APPLICANT: Barr, Philip
; APPLICANT: Gibson, Helen
; TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMODIUM
; TITLE OF INVENTION: FALCIPARUM VACCINE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Davis Hoxie Faithfull Hapgood
; STREET: 45 Rockefeller Pl.
; CITY: New York
; STATE: N.Y.
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,705
; FILING DATE: 14-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jacobs, Seth H
; REGISTRATION NUMBER: 32140
```


RESULT 5
US-08-290-919-4
; Sequence 4, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARRY & CUSHMAN, I. L. P.

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-290-919-12

Query Match 12.2%; Score 277; DB 1; Length 48;
Best Local Similarity 97.9%; Pred. No. 2.1e-14;
Matches 47; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 338 NISOHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNT 385
|||||
DB 1 NISOHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNT 48
|||||

RESULT 8

US-09-513-442-2
Sequence 2, Application US/09513442

Patent No. 6387664

GENERAL INFORMATION:

APPLICANT: Ikemoto, Mitsushi

TITLE OF INVENTION: SPARC Fusion Protein and a Process for Producing the

TITLE OF INVENTION: Same

FILE REFERENCE: HIRAKI-04218

CURRENT APPLICATION NUMBER: US/09/513,442

CURRENT FILING DATE: 2000-02-25

NUMBER OF SEQ ID NOS: 4

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 2

LENGTH: 467

TYPE: PRT

ORGANISM: Mus musculus

US-09-513-442-2

Query Match 12.1%; Score 275; DB 4; Length 467;
Best Local Similarity 98.0%; Pred. No. 6e-13;
Matches 50; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAARKFERQHMDSPDLGTDKAMADIGSI 51
|||||
DB 116 MHHHHHSSGLVPRGSGMKETAARKFERQHMDSPDLGTDKAMADIGSM 166
|||||

RESULT 9

US-08-290-919-2

Sequence 2, Application US/08290919

Patent No. 5720959

GENERAL INFORMATION:

APPLICANT: HOLDER, ANTHONY A.

APPLICANT: BLACKMAN, MICHAEL J.

APPLICANT: CHAPPEL, JONATHAN A.

TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA

TITLE OF INVENTION: VACCINE

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.

STREET: 1100 NEW YORK AVENUE, N.W.

CITY: WASHINGTON

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005-3918

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290,919

FILING DATE: 04-OCT-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9203821.5

FILING DATE: 22-FEB-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB93/00367
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.

REGISTRATION NUMBER: 16,773

REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC/8

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 861-3000

TELEFAX: (202) 822-0944

TELEX: 6714627 CUSH

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 48 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FEATURE:

NAME/KEY: Modified-site

LOCATION: 1

OTHER INFORMATION: /label= X

OTHER INFORMATION: /note= "X" = M and N, or "N"

US-08-290-919-2

Query Match 12.0%; Score 274; DB 1; Length 48;
Best Local Similarity 100.0%; Pred. No. 3.6e-14;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 339 ISOHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNT 385
|||||
DB 2 ISOHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNT 48
|||||

RESULT 10

US-09-741-243C-4

Sequence 4, Application US/09741243C

Patent No. 6399352

GENERAL INFORMATION:

APPLICANT: Crawford Jr., John Milton

APPLICANT: Rice, John

APPLICANT: Sevala, Veeresh

APPLICANT: Stewart, Sandy

TITLE OF INVENTION: NOVEL PLANT PORPHOBILINOGEN AND FUSION

TITLE OF INVENTION: PROTEIN THEREOF

FILE REFERENCE: 20220S

CURRENT APPLICATION NUMBER: US/09/741,243C

CURRENT FILING DATE: 2000-12-19

PRIOR APPLICATION NUMBER: 60/171,785

PRIOR FILING DATE: 1999-12-22

NUMBER OF SEQ ID NOS: 8

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4

LENGTH: 167

TYPE: PRT

ORGANISM: Unknown

FEATURE:

OTHER INFORMATION: Thioedoxin functional fragment

US-09-741-243C-4

Query Match 12.0%; Score 274; DB 4; Length 167;
Best Local Similarity 100.0%; Pred. No. 1.8e-13;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAARKFERQHMDSPDLGTDKAMADIGS 50
|||||
DB 116 MHHHHHSSGLVPRGSGMKETAARKFERQHMDSPDLGTDKAMADIGS 165
|||||

RESULT 11

US-09-741-243C-2

Sequence 2, Application US/09741243C

Patent No. 6399352

GENERAL INFORMATION:

APPLICANT: Crawford Jr., John Milton
APPLICANT: Rice, John
APPLICANT: Sevala, Veeresh
APPLICANT: Stewart, Sandy
TITLE OF INVENTION: NOVEL PLANT PORPHOBILINOGEN AND FUSION
TITLE OF INVENTION: PROTEIN THEREOF
FILE REFERENCE: 2022US
CURRENT APPLICATION NUMBER: US/09/741.243C
CURRENT FILING DATE: 2000-12-19
PRIOR APPLICATION NUMBER: 60/171,785
PRIOR FILING DATE: 1999-12-22
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 551
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: A plant thioresdoxin-porphobilinogen
OTHER INFORMATION: synthase fusion protein
US-09-741-243C-2

Query Match 12.0%; Score 274; DB 4; Length 551;
Best Local Similarity 100.0%; Pred. No. 8.8e-13;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRSGMKETAAAKFERQHMDSPDLGTDGDDKAMADIGS 50
|||||
DB 116 MHHHHHSSGLVPRSGMKETAAAKFERQHMDSPDLGTDGDDKAMADIGS 165

RESULT 12

US-08-290-919-3
Sequence 3, Application US/08290919
Patent No. 5720959
GENERAL INFORMATION:
APPLICANT: HOLDER, ANTHONY A.
APPLICANT: BLACKMAN, MICHAEL J.
APPLICANT: CHAPPEL, JONATHAN A.
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290.919
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9203821.5
FILING DATE: 22-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/00367
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 212242/HCM/NJL/6BC8/
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 53 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label- X
OTHER INFORMATION: /note= "X" - M and N, or N"
US-08-290-919-3

Query Match 12.0%; Score 272; DB 1; Length 53;
Best Local Similarity 94.0%; Pred. No. 5.8e-14;
Matches 47; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 302 PNPTCNENNGCDATCTEEDSGSRKKITCECTKPDSPFLDGFICSS 431
|||||
DB 2 PNPTCNENNGCDADAKCTEEDSGSGKKITCECTKPDSPFLDGFICSS 51

RESULT 13

US-08-290-919-1
Sequence 1, Application US/08290919
Patent No. 5720959
GENERAL INFORMATION:
APPLICANT: HOLDER, ANTHONY A.
APPLICANT: BLACKMAN, MICHAEL J.
APPLICANT: CHAPPEL, JONATHAN A.
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290.919
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9203821.5
FILING DATE: 22-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/00367
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 212242/HCM/NJL/6BC8/
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1


```

; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/09/340.479
; FILING DATE:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/946.475
; FILING DATE: 08-OCT-1997
; APPLICATION NUMBER: 60/027.973
; FILING DATE: 08-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50560
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-340-479-9

Query Match 6.6%; Score 151; DB 4; Length 711;
Best Local Similarity 22.9%; Pred. No. 0.0021;
Matches 68; Conservative 39; Mismatches 80; Indels 110; Gaps 13;

Qy 2 HHHHHSSGLVPRGSM-----KETAARKFERQHMDSPDLGTDGDDDDKAMADIGSIE 52
Db 5 HHHHHSSGLVPRGSMHTLADNLVIVSPAKATIEKYLGR-----KYKVIASMGHVR 57

Qy 53 G--RGTMAISVTMDNILSGFENEYDVIYKPLAGVYRSLKKOIE--KNIFTN----- 101
Db 58 DLPRQMGVD--TEDN-----YEPKY--ITIRKGPVVKELKKHAKAKNVFLASDPDRGE 110

Qy 102 -----LNLNDLNSRLKKRYFLDVLESIDLMOFKHISNEYIIEDSPKLLNSEQKN 152
Db 111 ATAWHLKILEDSKENRVVFNITKDAVKESFKNPREIEMN-----LVDAQOAR 161

Qy 153 TLLKS-----YKIKESV-----ENDI----- 169
Db 162 RILDLRLVGYNTSPVLWKVKKGLSAGRVQSVALLRLVIDRENEIRNFKPEYWTIEGEFRY 221

Qy 170 ---KFAQEGISV-----YEVKLYAKYKDDLESIKKVIKEKEKFPSSPPTT 211
Db 222 KKSFNKFLHYKKNKPKFLTKKDKVEKITAALDGDGDFEITNTVKKEKTRNANPPTT 278

RESULT 23
US-09-134-001C-5178
; Sequence 5178, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674

Qy 86 YRSLKKQIE---KNIFTNLDNLNLDLNSRLKKRYFLDVLE--SDLMQFKHISNEYIIE 140
Db 6361 YNNALKQAEDLIINNSSNPNLAQDITNA--LNNIKQAQNLHGAQKLDQDN--TTNQAI-- 6416

Qy 141 DSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYEVKLYAKYKDDLESIKKVIKEE 200
Db 6417 GNLNHLNQPQDALIQ-----AINGATSRDQVAEKLK--EAEALDEAMKQL 6460

Qy 201 KEKFPSSPPTTPPSPAKTDQEKESKFLPFLTNITLNNLVN-----KIDDYLINLKA 254
Db 6461 EDQVNDQDQISNSPFFINEDSDKQTKYNDKQAKEIINQTSNPLDKQKLTADTLQNKD 6520
; NUMBER OF SEQ ID NOS: 5674
```

```

; SEQ ID NO 5178
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5178

Query Match 6.1%; Score 139.5; DB 4; Length 1010;
Best Local Similarity 22.2%; Pred. No. 0.024;
Matches 71; Conservative 58; Mismatches 112; Indels 79; Gaps 14;

Qy 86 YRSLKKQIEKNIFTNLDNLNLDLNSRLKKRYFLDVLESIDLMOFKHISNEY--IIED 141
Db 348 YHNEIKGFQKOLEHLSTRENEITQFQYLEKNQVFNOLD-----KLISSYQOKPVTEE 401

Qy 142 SFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYEVKLYAKYKDDLESIKKVIKEE 201
Db 402 EIKRLYSEVNDLITKKEELTKEMNNKNKDF-----IIEHYTEIYKLKKIIDE-- 450

Qy 202 KEKFPSSPPTTPPSPAKTDQEKESKFLPFLTNITFLY-----NVLNKKIDDYLINLKA 254
Db 451 -----SERQKDEKLFQKLDKSSYLSKLKKEQLENEISSITNIDA 494

Qy 255 KIDCNVEKDEAHVKITKLSOLKAIDDKIDLFKNP-----YDPEATKKLINDDTKKDM 307
Db 495 TLIDUNDKKD---FVNEIKSAMSIGDTCPCGNEIHSLSGERHIDFESTAQ-----KNNK 544

Qy 308 LGKLSTGLVQNFNPNTIISKLTIEGKFQDM-----LNISOHCQVKKQCPENSGCFRHLDE 361
Db 545 IKRLSKKV--KIRDELIK--IETRIEELNHNRELNFEKQE--KKDISELQKQLNHLNQ 598

Qy 362 REECKLLNYKQEGDKCVEN 381
Db 599 -----LKDEQQSINKLVEN 612

RESULT 24
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 5.8%; Score 133; DB 4; Length 10182;
Best Local Similarity 24.2%; Pred. No. 1.5;
Matches 68; Conservative 49; Mismatches 126; Indels 38; Gaps 10;

Qy 86 YRSLKKQIE---KNIFTNLDNLNLDLNSRLKKRYFLDVLE--SDLMQFKHISNEYIIE 140
Db 6361 YNNALKQAEDLIINNSSNPNLAQDITNA--LNNIKQAQNLHGAQKLDQDN--TTNQAI-- 6416

Qy 141 DSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYEVKLYAKYKDDLESIKKVIKEE 200
Db 6417 GNLNHLNQPQDALIQ-----AINGATSRDQVAEKLK--EAEALDEAMKQL 6460

Qy 201 KEKFPSSPPTTPPSPAKTDQEKESKFLPFLTNITLNNLVN-----KIDDYLINLKA 254
Db 6461 EDQVNDQDQISNSPFFINEDSDKQTKYNDKQAKEIINQTSNPLDKQKLTADTLQNKD 6520
; NUMBER OF SEQ ID NOS: 5674
```


Matches	78;	Conservative	65;	Mismatches	135;	Indels	64;	Gaps
Qy	75	DVIYKPLAGVYSRLKKOIKNIFF-----NLNLDILNSLRKKRYFLDVL--- <td>127</td> <td>I : :</td> <td>I I I : :</td> <td>I : :</td> <td>I : :</td> <td>I : :</td>	127	I : :	I I I : :	I : :	I : :	I : :
Db	185	DAFHPPTHTSLRS-RDINRIELLKKYPNNIIDYSISKSNSSFIHLHMIISNM	243	I : :	I I I : :	I : :	I : :	I : :
Qy	128	OFKHISSNEI--IEDSFKLLNSEQ-KNTLTKSVKY----IKESVENDIKFAQGSIYYE	180	I : :	I : :	I : :	I : :	I : :
Db	244	PALIPSYNDISTVTVDKRDLJRMYGICVMFSYDNIMIDLESDDSDYIFIEKNISIYD	303	I : :	I : :	I : :	I : :	I : :
Qy	181	KVLAKYKDDESIAKVKEEKFPSSPPTTPPSPAKTDECKESKFLEFLTNETLYNN	240	I : :	I : :	I : :	I : :	I : :
Db	304	--VKCRDFAMIRDKVKRENRILT-----KCEDIIRIKLFSE-----KNRIND	346	I : :	I : :	I : :	I : :	I : :
Qy	241	LVNKIIDDYLNIKAKINCDCNVKEDAEHVKITKLSDLKAIDDKIDLFRNPYDFAEAKKLN	300	I : :	I I I : :	I : :	I : :	I : :
Db	347	ENNKVEEVLIHD-----NVSKNKK---LSLSDISSLMDQFRL--NPCTIRNI--LLS	392	I : :	I I I : :	I : :	I : :	I : :
Qy	301	DDTKKOMGLKLLSTGLLVNF-----PWITISKLEGKF-QDMLNISHQCVK--KQCPE	352	I : :	I : :	I : :	I : :	I : :
Db	393	SATIK---SKLLAURLAVKMCKCYSLTNVSYYKTIKGIVMDEVDTYSTNILKYHKQLDYK	449	I : :	I : :	I : :	I : :	I : :
Qy	353	SGCFRHLDERECEC-----LLNYKOEGDKVCVPENTCNE	388	I : :	I : :	I : :	I : :	I : :
Db	450	MSTEFYRKDIKSCACSICSDSITHHIYTETTSCL-NYKSTDND	490	I : :	I : :	I : :	I : :	I : :
RESULT 31								
US-08-480-640A-193								
:	: Sequence 193,	Application US/08480640A						
:	: Patent No.	6033904						
:	: GENERAL INFORMATION:							
:	: APPLICANT:	Cochran, Mark D.						
:	: APPLICANT:	Junker, David E.						
:	: TITLE OF INVENTION:	Recombinant Swinepox Virus						
:	: NUMBER OF SEQUENCES:	225						
:	: CORRESPONDENCE ADDRESS:							
:	: ADDRESSEE:	John P. White						
:	: STREET:	1185 Avenue of the Americas						
:	: CITY:	New York						
:	: STATE:	New York						
:	: COUNTRY:	USA						
:	: ZIP:	10036						
:	: COMPUTER READABLE FORM:							
:	: MEDIUM TYPE:	Floppy disk						
:	: COMPUTER:	IBM PC compatible						
:	: OPERATING SYSTEM:	PC-DOS/MS-DOS						
:	: SOFTWARE:	Patentin Release #1.0, Version #1.25						
:	: CURRENT APPLICATION DATA:							
:	: APPLICATION NUMBER:	US/08/480,640A						
:	: FILING DATE:	07-JUN-1995						
:	: CLASSIFICATION:	A24						
:	: ATTORNEY/AGENT INFORMATION:							
:	: NAME:	White, John P						
:	: REGISTRATION NUMBER:	28,678						
:	: TELECOMMUNICATION INFORMATION:							
:	: TELEPHONE:	(212) 278-0400						
:	: TELEFAX:	(212) 391-0525						
:	: INFORMATION FOR SEQ ID NO: 193:							
:	: SEQUENCE CHARACTERISTICS:							
:	: LENGTH:	677 amino acids						
:	: TYPE:	amino acid						
:	: TOPOLOGY:	linear						
:	: MOLECULE TYPE:	protein						
US-08-480-640A-193								
Query Match 5.5%; Score 126; DB 3; Length 677;								
Best Local Similarity 22.8%; Pred. No. 0.15;								
Matches	78;	Conservative	65;	Mismatches	135;	Indels	64;	Gaps
Qy	75	DVIYKPLAGVYSRLKKOIKNIFF-----NLNLDILNSLRKKRYFLDVL--- <td>127</td> <td>I : :</td> <td>I I I : :</td> <td>I : :</td> <td>I : :</td> <td>I : :</td>	127	I : :	I I I : :	I : :	I : :	I : :
Db	185	DAFHPPTHTSLRS-RDINRIELLKKYPNNIIDYSISKSNSSFIHLHMIISNM	243	I : :	I I I : :	I : :	I : :	I : :


```

; SOFTWARE: PatentIn Release #1.0, Version #1.25 .

```

```

; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,237A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 677 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-488-237A-193

```

```

Query Match          5.5%; Score 126; DB 4; Length 677;
Best Local Similarity 22.8%; Pred. No. 0.15;
Matches 78; Conservative 65; Mismatches 135; Indels 64; Gaps 19;

QY 75 DVIYKPLAGVYRSKQKQIEKNIFTF----NLNLDILNSRLKKRKYFLDVLE---SDLM 127
DB 185 DAFHPPHSHIRSL-RDINRIELKKYPNNNIIDYSDSIKSNSSFIHLHMIISNMF 243

QY 128 OFKHSSNEYI--IEDSFKLLNSEQ-KNTLKSRYK----IKESVENDIKFAQEGISYVE 180
DB 244 PAIIPSVNDFSTVVYDKORLINMYGKCVAMPSYDINMIDLESDDSDYFIEKNISYD 303

QY 181 KVLAKYKDDLESIKKVIKEKEKFPSSPTTPPSPAKTDEQKESKFLPFLTNITLYNN 240
DB 304 ---VKCRDFANNIRDKVKREKNRIILT-----KCEDIIRYIKLFS-----KNRIND 346

QY 353 SGCPRHLDERECKC-----LLNYKQEGDKCVENPNTCNE 388
DB 450 MSTFEYKRDIKSKCKSCISDSITHIYETTSCI-NYKSTDND 490

```

```

RESULT 37
US-08-375-992A-115
; Sequence 115, Application US/08375992A
; Patent No. 6328975
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 220
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,992A
; FILING DATE: Herewith
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 677 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Swinepox virus
; STRAIN: Kasza
; INDIVIDUAL ISOLATE: S-SPV-001
; IMMEDIATE SOURCE:
; CLONE: 515-85.1
; POSITION IN GENOME:
; MAP POSITION: -23.2
; UNITS: %G
; US-08-375-992A-115

```

```

Query Match          5.5%; Score 126; DB 4; Length 677;
Best Local Similarity 22.8%; Pred. No. 0.15;
Matches 78; Conservative 65; Mismatches 135; Indels 64; Gaps 19;

QY 75 DVIYKPLAGVYRSKQKQIEKNIFTF----NLNLDILNSRLKKRKYFLDVLE---SDLM 127
DB 185 DAFHPPHSHIRSL-RDINRIELKKYPNNNIIDYSDSIKSNSSFIHLHMIISNMF 243

QY 128 OFKHSSNEYI--IEDSFKLLNSEQ-KNTLKSRYK----IKESVENDIKFAQEGISYVE 180
DB 244 PAIIPSVNDFSTVVYDKORLINMYGKCVAMPSYDINMIDLESDDSDYFIEKNISYD 303

QY 181 KVLAKYKDDLESIKKVIKEKEKFPSSPTTPPSPAKTDEQKESKFLPFLTNITLYNN 240
DB 304 ---VKCRDFANNIRDKVKREKNRIILT-----KCEDIIRYIKLFS-----KNRIND 346

QY 241 LVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIKDDKIDLFKNPYDEAIKKLIN 300
DB 347 ENNKVEEVLHID-----NYSKNNK-----LSLSDISSLMQDFRL--NPCTIRNI--LLS 392

QY 301 DDTKDDMLGKLLSTGLVONF-----PNTIISKLEGEKQF-QDMLNISQHCQVK--KQCPEN 352
DB 393 SATIK---SKLLALRAVNWKCYSLTNVSMYKKIKGVIMDMVDYISTNILKYHKQLYDK 449

QY 353 SGCPRHLDERECKC-----LLNYKQEGDKCVENPNTCNE 388
DB 450 MSTFEYKRDIKSKCKSCISDSITHIYETTSCI-NYKSTDND 490

```

```

RESULT 38
US-08-375-992A-193
; Sequence 193, Application US/08375992A
; Patent No. 6328975
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 220
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas

```

CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/375,992A

FILING DATE: Herewith

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: White, John P

REGISTRATION NUMBER: 28,678

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 193:

SEQUENCE CHARACTERISTICS:

LENGTH: 677 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-375-992A-193

Query Match 5.5%; Score 126; DB 4; Length 677;

Best Local Similarity 22.8%; Pred. No. 0.15;

Matches 78; Conservative 65; Mismatches 135; Indels 64; Gaps 19;

QY 75 DVIYKPLAGVYSLKQTEKNTFTF-----NLNLDILNSRLKRYFLDVLVLE---SDLM 127

DB 185 DAFHIPPTHSLSRL-RDINRIIELLKYPNNIIDIYISDSIKSSSFHILHNIISNMF 243

QY 128 OFKHISSEYI--IEDSFLLNSEQ-KNTLLKSYKY-----IKESVENDIKFAEGISYE 180

DB 244 PAIIPSVNDFISTVVDKRLNINMGTKVAMFSYDINMIDLESDDSDYIFIEKNISYD 303

QY 181 KVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAKTOEQKESKFLPFLTNIELYNN 240

DB 304 ---VKCRDFANMRDVKREKNRLIT-----KCDIIRYIKLFS-----KRNID 346

QY 241 LVNKIDYILINKAKINDCNVEKDEAHVITKLSDLKAIDDKTDLKPNVDFEAIKKLIN 300

DB 347 ENNKVEEVLTHID-----NVSKNNK-----LSLSDISSLMQFRL--NPCTIKNI--LLS 392

QY 301 DPTKKDMLKLLSTGLVONF-----PNTIISKLECKF-QDMLNISQHCVK--KQCPEN 352

DB 393 SATIK---SKLLALRAVKNKCYSLTNVSMYKKIKGVIVMDVYISTNLIKHYKQLYDK 449

QY 353 SGCFRHLDERECKC-----LLNYKQEGDKCVENPNPTCNE 388

DB 450 MSTFEYKRDIKSKCSCICSDSITHIYETTSICI-NYKSTDND 490

RESULT 39

US-08-287-959-1

Sequence 1, Application US/08287959

Patent No. 5639651

GENERAL INFORMATION:

APPLICANT: Weissbach, Lawrence

APPLICANT: Bernards, Andre

APPLICANT: Settlemann, Jeffrey

TITLE OF INVENTION: GAP-RELATED GENE

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESS: Fish & Richardson

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: U.S.A.

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/287,959

FILING DATE: August 9, 1994

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Clark, Paul C.

REGISTRATION NUMBER: 30,162

REFERENCE/DOCKET NUMBER: 00786/181001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1657 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-287-959-1

Query Match 5.5%; Score 125.5; DB 1; Length 1657;

Best Local Similarity 17.5%; Pred. No. 0.52;

Matches 69; Conservative 81; Mismatches 106; Indels 139; Gaps 16;

QY 75 DVIYKPLAGVYSLKQTEKNTFTFNLNLDILNSRL-----KKRYFLDVLVLESD---- 125

DB 808 EVYKIQSLARMHQA-RKRYRDRLOQYFRDHINDIKIOAFIRANKARDDYKTLINAEPPM 866

QY 126 --LMQFKHI--SSNEVIYEDSFLLNSEQNTLLKSVKYIKESVENDIKE----- 171

DB 867 VVYKRVHLLDQSDQDFQELDLKMKREEVITLIRS-----NQLENDLNLMDIKIGLLVK 922

QY 172 ----AQEGISYEVKVLAKYKDDLE-----SIKKVIKEKEKFPSPSP-----PTT 211

DB 923 NKITLQDVVSHSKLTAKNKEQLSDMMINKQGGKLSKREKLEAYOHLFYLLQTN 982

QY 212 PPSFATDQKESKESKELPFLTN-IETLYNNLVNKIDYLI-----NLKAKINDCN 260

DB 983 PTYLAKLIFQMPQNKSTKFMDSVIFTLYNASQREYLLLRFLKFTALOEIEKSKVDQIQ 1042

QY 261 --VEKDEAHVK-----ITKLSDLKAID--DKIDLEFN----- 288

DB 1043 EIVTGNPTVTKVVSFNRCARGONALRQILAPVVKEMDDKSLNKTDPVDYIKSVNQM 1102

QY 289 -----PYDFAIKKLLINDDTTKDM----- 307

DB 1103 ESQTGEASKLPYDVTPEQALAHAEVKTRLDSSIRNMRAVTDKLSAIVSSVDKIPYGMRF 1162

QY 308 LKLLSTGLVQNEPNT-----IISKLEKGFQD 335

DB 1163 IAKVLKDSLHKEPFDAGEDELLKILGNLLYRYNM 1197

RESULT 40

US-08-235-836C-74

Sequence 74, Application US/08235836C

Patent No. 6248562

GENERAL INFORMATION:

APPLICANT: Dunn, John J.

APPLICANT: Luft, Benjamin J.

TITLE OF INVENTION: No. 6248562el Chimeric Proteins Comprising

TITLE OF INVENTION: Borrelia Polypeptides and Uses Therefor

NUMBER OF SEQUENCES: 144

CORRESPONDENCE ADDRESS:

ADDRESS: Brookhaven National Laboratory

STREET: Upton

CITY: Upton

STATE: NY

COUNTRY: USA
ZIP: 11973
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/235,836C
FILING DATE: 29-APR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,191
FILING DATE: 01-11-93
ATTORNEY/AGENT INFORMATION:
NAME: Bogosian, Margaret C.
REGISTRATION NUMBER: 25,324
REFERENCE/DOCKET NUMBER: BNL93-28A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 282-7338
TELEFAX: (516) 282-3729
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 700 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-235-836C-74

Query Match 5.5%; Score 125; DB 4; Length 700;
Best Local Similarity 20.4%; Pred. No. 0.18; Indels 114; Gaps 16;
Matches 84; Conservative 61; Mismatches 153;

Qy	39	DDDKAMADIGSIEGRGTMAISVTMDNLSGFENEYDIVYKPLAGVYRSILKQKIEKNIF	98
Db	88	DDDKASVDIFSIGSKSELDLSILNRLITG-----YLMKSFYERSAELIAKAIT	139
Qy	99	TFNLMNDILNSRLKRYFDLVESDLMQPKHISSNEYIYIEDSFKLNSQKNTLLKSY	158
Db	140	YNA-----VYRGDLDYK-----EFYIEASLKSILTKEKAG-LSRVY	175
Qy	159	-----KYIKESVENDIKFAQEGISYERKVLAKYKDDLESI--KKVI----	197
Db	176	SQWAGKTOIFIPLKNILSGVESDI-----DIDSLVTDKVVAAALL	216
Qy	198	--KEEKEFPSPPTPPSPAKTDEQKESKFLPPLNIETLYNNLVNKIDYILINKAK	255
Db	217	SENEGVNFARDITDIQETHKADQKIDIE---LDNIHESDSNITETIENLRDLEKA	272
Qy	256	INDCNVEKDEAHVKITKLS-----DLKAITD-----DKIDLKFNPYDF-----EAIKKLIN	300
Db	273	TDEEHKEIESQVDAKKKQKELDKKAIDLDKAQOKLDFDAEDNLDIQDVTVREKLOENIN	332
Qy	301	DDTKKDLMLGKLLSTGLVGNFPNTIISKLEGKFO--DMLNISQHCYKQKCFENSGCFRH	358
Db	333	ETNKEKNLPK-----PGDVSSPKVQQLQIKESLEDLQEQ-LKETGDEN-----	375
Qy	359	LDERECKCLLNKQEGDKVCNPNPTCNENNGCGDADATCTEEDSGSSRRK	410
Db	376	--QKREIEKQIEKSKDEKLKSKDDRKASKDGKALDLDRELSKASKEKSK	425

Search completed: May 12, 2003, 10:20:56
Job time : 43 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 19, 2003, 15:54:27 ; Search time 131 Seconds
(without alignments)
4243.995 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 2275

Sequence: 1 MHHHHHSSGLVPRGSMKE.....TCECTKPSYLFDFGIFCSS 431

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 810007 seqs, 644969091 residues

Total number of hits satisfying chosen parameters: 1620014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/usPTO_spool/US10057531/runat_12052003_091033_21772/app.query.fasta_1.583
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10057531@cgn_1_1_96_grunat_12052003_091033_21772
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1928	84.7	1235	9 US-10-098-514-13	Sequence 13, Appl
2	1911	84.0	1149	9 US-10-098-514-3	Sequence 3, Appl
3	1900	83.5	1149	9 US-10-098-514-1	Sequence 1, Appl
4	1175	51.6	5917	9 US-10-087-464-9	Sequence 9, Appl

5	1165.5	51.2	1131	9 US-10-087-464-58	Sequence 58, Appl
6	1137	51.2	1137	9 US-10-087-464-54	Sequence 54, Appl
7	1161.5	51.1	1065	12 US-10-082-018-1	Sequence 1, Appl
8	1161.5	51.1	1088	12 US-10-082-018-2	Sequence 2, Appl
9	1154.5	50.7	1142	12 US-10-082-018-8	Sequence 8, Appl
10	527	23.2	330	10 US-09-134-333-9	Sequence 9, Appl
11	527	23.2	342	10 US-09-134-333-6	Sequence 6, Appl
12	527	23.2	354	10 US-09-134-333-4	Sequence 4, Appl
13	523	23.0	279	10 US-09-134-333-3	Sequence 3, Appl
14	523	23.0	291	10 US-09-134-333-1	Sequence 1, Appl
15	523	23.0	387	10 US-09-134-333-7	Sequence 7, Appl
16	518	22.8	343	9 US-10-087-464-59	Sequence 59, Appl
17	306.5	13.5	1551	10 US-09-804-626-3	Sequence 3, Appl
18	304.5	13.4	1557	10 US-09-804-626-1	Sequence 1, Appl
19	278.5	11.5	2472	9 US-10-046-583A-6	Sequence 6, Appl
20	262.5	11.5	1055	9 US-10-012-896-1010	Sequence 1010, Ap
21	168	7.4	150	9 US-10-098-514-15	Sequence 15, Appl
22	144	6.3	2988	10 US-09-815-242-4196	Sequence 4196, Ap
23	144	6.3	3030	10 US-09-815-242-8043	Sequence 8043, Ap
24	140.5	6.2	2943	10 US-09-888-615-6	Sequence 6, Appl
25	137	6.0	3057	10 US-09-974-300-2660	Sequence 2660, Ap
26	132	5.8	2760	9 US-09-893-519A-146	Sequence 146, App
27	131.5	5.8	2847	9 US-10-267-311-20	Sequence 20, Appl
28	131	5.8	2853	10 US-09-764-864-373	Sequence 373, App
29	129	5.7	2341	10 US-09-881-752A-147	Sequence 147, App
30	129	5.7	6386	9 US-10-098-841-40	Sequence 40, Appl
31	128	5.6	1176	9 US-10-046-935-2238	Sequence 2238, Ap
32	128	5.6	1176	9 US-10-146-502-2238	Sequence 2238, Ap
33	127	5.6	4198	10 US-09-964-824A-263	Sequence 263, App
34	127	5.6	4567	12 US-10-044-090-102	Sequence 102, App
35	126.5	5.6	6773	10 US-09-864-864-336	Sequence 336, App
36	126	5.5	1678	10 US-09-974-300-2643	Sequence 2643, Ap
37	125.5	5.5	7573	10 US-09-880-107-2195	Sequence 2195, Ap
38	125	5.5	142	9 US-10-098-514-16	Sequence 16, Appl
39	123	5.4	2405	9 US-09-298-523B-70	Sequence 70, Appl
40	123	5.4	3463	9 US-09-298-523B-59	Sequence 59, Appl
41	122.5	5.4	8133	9 US-10-011-366-5	Sequence 5, Appl
42	121.5	5.3	12313	9 US-10-171-311-7	Sequence 7, Appl
43	121.5	5.3	12337	9 US-10-171-311-5	Sequence 5, Appl
44	121.5	5.3	12438	9 US-10-171-311-3	Sequence 3, Appl
45	121.5	5.3	12462	9 US-10-171-311-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-10-098-514-13
; Sequence 13, Application US/10098514
; Publication No. US20020194648A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra P
; APPLICANT: Christopher, David A
; APPLICANT: Vine, Benjamin
; APPLICANT: Su, Wei-wen
; APPLICANT: Bugos, Robert
; TITLE OF INVENTION: PLASMIDIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PRO
; FILE REFERENCE: A-71339/RET/TAL/NBC
; CURRENT APPLICATION NUMBER: US/10/098,514
; CURRENT FILING DATE: 2002-08-06
; PRIOR APPLICATION NUMBER: US 09/500,376
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: US 60/274,599
; PRIOR FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 1235
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
US-10-098-514-13

Alignment Scores:

Pred. No.:	1.9e-169	Length:	1335
Score:	1928.00	Matches:	368
Percent Similarity:	98.67%	Conservative:	2
Best Local Similarity:	98.13%	Mismatches:	5
Query Match:	84.75%	Indels:	0
DB:	9	Gaps:	0

US-10-057-531A-2 (1-431) x US-10-098-514-13 (1-1235)

QY	56	ThrMetAlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluThrAsp	75
DB	73	ACAGACGCGATCTGTACTATGGACAACATCTCTCAGTGGCTTCGAGAACGAGTACGAC	132
QY	76	ValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLys	95
DB	133	GTAACTCTACCTAAAGCCCTTGC GGGTGCTACCGTTCATTGAGAAACAGATAGAAAG	192
QY	96	AsnIlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArg	115
DB	193	AATATTTTTCACGTTCAACCTCAACCTTAAATAGACATCTCTCAACTCGCGCCTCAAGAAGCGA	252
QY	116	LysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsn	135
DB	253	AAATACTCTCTCGACGTGTGGAAATCCGACCTTATCGAATTTAAGCACATTTAGCTCTCAAC	312
QY	136	GluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeu	155
DB	313	GAGTACATCATAGCAGCAGCTTCAAGCTCTGAATTCAGACAGAAACACCTCCTCTA	372
QY	156	LysSerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGly	175
DB	373	AAGTCTCTACAAATACATTAGGAGTCTGTGTGAGAACGACATCAAGTTCGCCCGAAGGA	432
QY	176	IleSerTyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLys	195
DB	433	ATTAGCTACTATCAGAAAGTCTCTGGCTTAATACNAGACGACCTTGGAAAGCATTTAGAAAG	492
QY	196	ValIleLysGluLysGluLysPheProSerSerProThrThrProProSerPro	215
DB	493	GTAAATCAAGAAGAGAGGAAAAAGTTTCGAGCTCTCCACCCACAACCTCCCCCATCGCCT	552
QY	216	AlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGlu	235
DB	553	GCAAGACCCGACGACGAGAAAAAAGAAAGTAAGTTCCTTCATCTCTCACCACCATCGAA	612
QY	236	ThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLys	255
DB	613	ACTCTATATAACAACCTGGTGGAACAAGATTGATGACTACTTAATCAACTTGAAGCGCAAA	672
QY	256	IleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAsp	275
DB	673	ATTAATGACTGTAACTGCGAAAAAGGATGAAGCCCGTAAAGATCACCAAGCTTTCCGAT	732
QY	276	LeuLysAlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIle	295
DB	733	CTCAAGCCATCCAGCAGTATGATGACCTGTTTAAGAACCCACACCATTTTCAGCCGCAATC	792
QY	296	LysLysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGly	315
DB	793	AAAAAGTTGATCAACGACGATACTAAGAAAGACATGCTTGGAAAACTGCTGTGCACAGCG	852
QY	316	LeuValGlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAsp	335
DB	853	TTGGTCCAAACATTCGCCAACACCATATTAAGCAAGCTGATCGAAGAAAGTTTCAGGAT	912
QY	336	MetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCys	355
DB	913	ATGCTGNACATCTCTCAGCATCAATCGGTGCAAGACGAATGTCCTCCGAGAAATTCAGGTTGC	972
QY	356	PheArgHisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGly	375

```
QY 141 AspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyr 160
Db 256 GACAGCTTCAAGCTCTTGAATTCAGAACAGAACACACCTCTCTAAGTCTCAAAATAC 315
QY 161 IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluClyIleSerTyrTyrGlu 180
Db 316 ATTAAAGAGTCTGTTGAGAACGACATCAAGTTCGCCAGGAGGAATAGCTACTATGAG 375
QY 181 LysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluGlu 200
Db 376 AAAGTCTCGCTAAATACAGGAGGACTTGGAAAGGATTAAGAAAGTAATCAAGAGAG 435
QY 201 LysGluLysPheProSerProProThrThrProProSerProAlaLysThrAspGlu 220
Db 436 AAGGAAAAGTTTCGAGCTCTCCACCCACACACTCCCCCTGCAAGACCGACGAG 495
QY 221 GlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsn 240
Db 496 CAGAAAAAGAAAGTAAGTTCCTTCCATTCCTCACCAACATCGAAACTCTATATAACAC 555
QY 241 LeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsn 260
Db 556 CTGGTGAACAGATGATGACTACTTAATCAACTTGAAGCGGAAATTAATGACTGTAAAC 615
QY 261 ValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAsp 280
Db 616 GTCGAAAAGGATCAAGCCACGTTAAGATCACCAAGCTTCCGATCTCAAAAGCCATCGAC 675
QY 281 AspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsn 300
Db 676 GATAAGATTGACCTGTTTAAGAACCAACACAGATTCGACGCAATCAAAAGTTGATCAAC 735
QY 301 AspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPhe 320
Db 736 GAGGACTACTAAGAAACACATGCTTGAAAACTGCTGCGACAGGCTTGGTCCAAAACCTC 795
QY 321 ProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSer 340
Db 796 CCGAACACCATTAATAAGCAAGCTGATCGAAGGAAAGTTTCAGGATATGCTGAACATCTCT 855
QY 341 GlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAsp 360
Db 856 GAGCATCAATGCGTGAAGAACCAATGTCGCGAGAATTCAGGTTGCTTCGCGCACTTAGAC 915
QY 361 GluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGlu 380
Db 916 GAAAGGAGGAGATGTAATGCGCTGCTGAATATAACAGAGAGAGACAAAGTCCGTAGAG 975
QY 381 AsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThr 400
Db 976 AATCCCTAACCCACCCCTGTAACGAAAAATAACGGTGGCTGCGATGCTGACGCTAAGTGTACC 1035
QY 401 GluLysAspSerClySerArgLysLysIleThrCysGluCysThrLysProAspSer 420
Db 1036 GAGGAGACAGCGGTTCCCAATGCGAAGAAATAAATTCGGAATGCACGAGCCCGATAGT 1095
QY 421 TyrProLeuPheAspGlyIlePheCysSer 430
Db 1096 TACCCTCTCTTCGACGGTATCTTCTGCTCC 1125
```

RESULT 3

```
US-10-098-514-1
; Sequence 1, Application US/10098514
; Publication No. US20020194648A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra P
; APPLICANT: Christopher, David A
; APPLICANT: Vine, Benjamin
; APPLICANT: Su, Wei-wen
; APPLICANT: Bugos, Robert
; TITLE OF INVENTION: PLASMIDIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PRODUCE
; TITLE OF INVENTION: TRANSGENIC PLANTS
; FILE REFERENCE: A-71339/RET/TAL/NBC
```

```
; CURRENT APPLICATION NUMBER: US/10/098,514
; CURRENT FILING DATE: 2002-08-06
; PRIOR APPLICATION NUMBER: US 09/500,376
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: US 60/274,599
; PRIOR FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1149
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (1)..(24)
; OTHER INFORMATION:
US-10-098-514-1

Alignment Scores:
Pred. No.: 6,74e-167 Length: 1149
Score: 1300.00 Matches: 361
Percent Similarity: 98.38% Conservative: 4
Best Local Similarity: 97.30% Mismatches: 6
Query Match: 83.52% Indels: 0
DB: 9 Gaps: 0
```

US-10-057-531A-2 (1-431) x US-10-098-514-1 (1-1149)

```
QY 60 ServValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeu 79
Db 13 AACATCTCTCAGTCACAACTCCTCAGTGGCTTCGAGAACGAGTACGACGTAATCTACCTA 72
QY 80 LysProLeuAlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIlePheThr 99
Db 73 AACCCCTTCGCGGTGCTTACCCTTCATTGAAGAAACAGATAGAAAATAATTTTCACG 132
QY 100 PheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeu 119
Db 133 TTCAACTCAACCTTAATGACATCTCCTCACTCGGCTCAAGAAAGCGAAAATCTCTCCTC 192
QY 120 AspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIle 139
Db 193 GACGTGTGGAATCCGACCTTATGCAATTCAGACACATTAGCTCTCAAGAGTACATCATA 252
QY 140 GluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLys 159
Db 253 GAGGACAGCTTCAAGCTCTTGAATTCAGAACAGAACACCCCTCTCTAAAGTCTCTACAAA 312
QY 160 TyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyr 179
Db 313 TACATTAAAGGAGTCTGTTGAGAACGACATCAAGTTCGCCAGGAGGAATTAGCTACTAT 372
QY 180 GluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGlu 199
Db 373 GAGAAAGTCTCGCTAAATACAAGGACGACTTGGAAAGCATTAAGAAAGGTAATCAAGAA 432
QY 200 GluLysGluLysPheProSerSerProProThrThrProProSerProAlaLysThrAsp 219
Db 433 GAGAAAGGAAAGTTTCGAGCTCTCCACCCCACTCCCTCCATCGCTGCAAAAGACGAC 492
QY 220 GluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsn 239
Db 493 GAGCAGAAAAAGAAAGTAAGTTCCTTCCATTCCTCACCACATCGAACTCTATATAAC 552
QY 240 AsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCys 259
Db 553 AACCTGGTGAACAGATTGATGACTACTTAATCAACTTGAAGCGGAAAAATTAATGACTGT 612
QY 260 AsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIle 279
Db 613 AACGTCGAAAAGGATGAAGCCACGTTAAGATCACCAAGCTTTCCGATCTCTCAAGCCATC 672
```

QY 280 AspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIle 299
 |||||
 DB 673 GACGATAAGATTGACCTGTTTAAAGAACCAACACGATTTTCGACCAATCAAAAAGTTGATC 732
 |||||
 QY 300 AsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsn 319
 |||||
 DB 733 AACGACGATACCTAAGAAAGACATGCTTGAAACCTGCTGTCACAGGCTTGGTCCAAAAC 792
 |||||
 QY 320 PheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeuAsnIle 339
 |||||
 DB 793 TTCCCGAACACCATTAAGCAAGCTGATCGAAGGAAGTTTCAGGATATGCTGCAACATC 852
 |||||
 QY 340 SerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeu 359
 |||||
 DB 853 TCTCAGCATCAATGCGTGAAGAAAGCAATGCTCCGAGAATTACAGTGGTTCGCCACATTA 912
 |||||
 QY 360 AspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysVal 379
 |||||
 DB 913 GACGAAGAGGAGGAGTAAATGCTGCTGCTGAATTAATAACAGGAAGGAGACANGTCCGTA 972
 |||||
 QY 380 GluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCys 399
 |||||
 DB 973 GAGAATCCTAACCCCAACCTGTAAACGAAATAACGGTGGCTGCGATGCTGACGCTAAGTGT 1032
 |||||
 QY 400 ThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLysProAsp 419
 |||||
 DB 1033 ACCGAGGAGGACAGCGGTTCGAATGGCAAGAAATAAATTCGAATGCGAAGGCCGAT 1092
 |||||
 QY 420 SerTyrProLeuPheAspGlyIlePheCysSer 430
 |||||
 DB 1093 AGTTACCTCTCTCGACGGTATCTTCGTCTCC 1125
 |||||

RESULT 4

US-10-087-464-9
 ; Sequence 9, Application US/10087464
 ; Publication No. US20030059436A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Chishtli, Athar
 ; APPLICANT: Oh, Steven
 ; APPLICANT: Liu, David
 ; APPLICANT: Goel, Vikas
 ; APPLICANT: Li, Xuerong
 ; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
 ; CURRENT APPLICATION NUMBER: US/10/087,464
 ; PRIOR FILING DATE: 2002-03-01
 ; PRIOR APPLICATION NUMBER: US 06/272,930
 ; NUMBER OF SEQ ID NOS: 59
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO: 9
 ; LENGTH: 5917
 ; TYPE: DNA
 ; ORGANISM: Plasmodium falciparum
 US-10-087-464-9

Alignment Scores:
 Pred. No.: 9,14e-99 Length: 5917
 Score: 1175.00 Matches: 226
 Percent Similarity: 71.43% Conservative: 59
 Best Local Similarity: 56.64% Mismatches: 90
 Query Match: 51.65% Indels: 24
 DB: 9 Gaps: 4

US-10-057-531A-2 (1-431) x US-10-087-464-9 (1-5917)

QY 34 ProAspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGly 53
 |||||
 DB 4147 CCTATTATTGGAGAAATCCGAGAAGAT-----TATGATGATTTAGGACAAAGTAGTAACA 4200
 |||||
 QY 54 ArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGlu 73
 |||||
 DB 5218 TGTGAATGTACTAAACCTGATCTTATCCACTTTTCATGGTATTTTCTCAGGTTC 5274
 |||||

DB 4201 GGAGAGCAGTAACCTCTTCCGTA---ATTGATAACATACTTTCTAAAAATGAAATGAA 4257
 QY 74 TyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIle 93
 |||||
 DB 4258 TATGAGGTTTATATATTTAAACCTTTACGAGGTGTTTATAGAGGTTTAAAAAACAATTA 4317
 |||||
 QY 94 GluLysAsnIlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLys 113
 |||||
 DB 4318 GAAATAACGTTTATCAGATTTAATGTTAATGTTAAGGATATTTTAAATTCACGATTAAT 4377
 |||||
 QY 114 LysArgLysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSer 133
 |||||
 DB 4378 AAACGTGAAATTTTCAAAAATGTTTAAATCAGATTTTAAATTCATATAAGATTTAAACA 4437
 |||||
 QY 134 SerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThr 153
 |||||
 DB 4438 TCAAGTAATTTATGTTGTCNAAGATCCATATAATTTCTTAATAAGAAAAAAGACATAAA 4497
 |||||
 QY 154 LeuLysSerTyrTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGln 173
 |||||
 DB 4498 TTCTTTAAGCAGTTATATAATTAAGGATTCATAGATACGATATAAATTTTGCATTAAT 4557
 |||||
 QY 174 GluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIle 193
 |||||
 DB 4558 GATGTTCTGGATATTATAAATATTATCCGAAAAATATAAATCAGATTTAGATTTCAATT 4617
 |||||
 QY 194 LysLysValIleLysGluGluLysGluLysPheProSerSerProThrThrProPro 213
 |||||
 DB 4618 AAAAAATATATC----- 4629
 QY 214 SerProAlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsn 233
 |||||
 DB 4630 -----AACGACAAAACAGGTGAAATAGAAATACCTTCCCTTTTAAACAAT 4677
 |||||
 QY 234 IleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLys 253
 |||||
 DB 4678 ATTGAGACCTTATATAAAGAGTTAATGATAAATGATTTATTTGTAATTCATTTAGAA 4737
 |||||
 QY 254 AlaLysIleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeu 273
 |||||
 DB 4738 GCAAAAGTTCTAAATTTATACATATACAGAAATCAAAAGTAGAAGTTTAAATAAACAATT 4797
 |||||
 QY 274 SerAspLeuLysAlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGlu 293
 |||||
 DB 4798 AATTACTTTAAACAATTTCAAGACAAATTTGCGAGATTTTAAAAAAAATAACAATTTCTGT 4857
 |||||
 QY 294 AlaIleLysLysLeuIleAsnAspThrLysLysAspMetLeuGlyLysLeuLeuSer 313
 |||||
 DB 4858 GGAATTGCTGATTTATCAACAGATTTAACCATAATACTTATTGACAAAGTTCTCTAGT 4917
 |||||
 QY 314 ThrGlyLeuVal---GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLys 332
 |||||
 DB 4918 ACAGGTATGTTTGTGAAATCTTCTGCTAAACCGTTTTATCTAATTTACTTGTGGAAC 4977
 |||||
 QY 333 PheGlnAspMetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsn 352
 |||||
 DB 4978 TTGCAAGGTATGTTTAAACATTTTCACACCAATTCGCTGCTGCTGCTGCTGCTGCTGCTG 5037
 |||||
 QY 353 SerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLys 372
 |||||
 DB 5038 TCTGGATGTTTCAGACATTTAGATGAAGAGAAGAAATGTAATGTTTATTAATTAACAAA 5097
 |||||
 QY 373 GlnGluGlyAspLysCysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGly 392
 |||||
 DB 5098 CAAGAAGGTGATAAATGTTGAAATCCCAATCTCTACTTCTAAGCAAAATAATGTTGGA 5157
 |||||
 QY 393 CysAspAlaAspAlaThrCysThrGluCysSerGlySerGlySerArgLysIleThr 412
 |||||
 DB 5158 TGTGATCAGATGCCAATGTTACCGAAGAGATTCAGGTAGCAACGGAAGAAATAACACA 5217
 |||||
 QY 413 CysGluCysThrLysProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
 |||||
 DB 5218 TGTGAATGTACTAAACCTGATCTTATCCACTTTTCATGGTATTTTCTCAGGTTC 5274
 |||||

Qy 63 MetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeu 82

```
Db 25 ATGATACATACTTCTAAATGAAATGAATGAGTTTATATATATAAACCCTTA 84
Qy 83 AlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIlePheThrPheAsnLeu 102
Db 85 GCAGCTGTTATAGAAGTTTAAAAAACAATAGAAATACGTTATGACATTTAATGTT 144
Qy 103 AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeu 122
Db 145 AATGTTAAGGATATTAAATTCAGGATTAAATAACGTCGAAATTTCAAAATCTTTTA 204
Qy 123 GluSerAspLeuMetGlnPheLysHisIleSerAsnGluTyrIleIleGluAspSer 142
Db 205 GAATCAGATTAAATCCATATAAGATTATACATCAAGTAATATGTTGTCAAGATCCA 264
Qy 143 PheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSerTyrLysTyrIleLys 162
Db 265 TATAAATTTCTTAATAAGAAAAAGAGATAAATCTTAAGCAGTTATATATATTAAAG 324
Qy 163 GluSerValGluAsnAspIleLysPheAlaGlnGluLysIleSerTyrTyrGluLysVal 182
Db 325 GATTCATAGATACGATATAAATTTGCAATGATGTTCTTGGATATATAAATAATTA 384
Qy 183 LeuAlaLysTyrLysAspLeuGluSerIleLysValIleLysGluLysGlu 202
Db 385 TCCGAAAAATATAAATCAGATTAGATTCAATTAATAAATAATATATC----- 429
Qy 203 LysPheProSerSerProProThrProProSerProAlaLysThrAspGluGlnLys 222
Db 430 -----AAGCAACAACAAGGT 444
Qy 223 LysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnLeuVal 242
Db 445 GAAATGAGAATACCTTTCCCTTTTAAACAATATTGAGACCTTATATAAACAATTAAT 504
Qy 243 AsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGlu 262
Db 505 GATAAATGATTATTATGTAATCATTTAGAACGAAAGTTCTAAATATTATACATATGAG 564
Qy 263 LysAspGluAlaHisValLysIleThrLysLeuSerAspLysAlaIleAspLys 282
Db 565 AAATCAAAACGTAGAAGTTAAATAAAGAACTTAATTAATTTAAACAATTTCAAGACAAA 624
Qy 283 IleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAsp 302
Db 625 TTGCAGATTTTAAAAAATAAACAATTTCTGTAATTTGCTGATTATATCAACAGATTAT 684
Qy 303 ThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuVal---GlnAsnPhePro 321
Db 685 AACCATAATACTATTGACAAAGTTCTTACTAGCAGTATGGTTTTTGAAATCTTGCT 744
Qy 322 AsnThrIleLeuSerLysLeuIleGluLysPheGlnAspMetLeuAsnIleSerGln 341
Db 745 AAAACCGTTTATCTAATTTACTTGAGAAACTTGCAAGGTATGTTAAACATTCACAA 804
Qy 342 HisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGlu 361
Db 805 CACCAATCGGTAAATAAACAATGTCACAAATTTCTGCGATGTTTCAGACATTTAGATGAA 864
Qy 362 ArgGluGluCysLysCysLeuAsnTyrLysGlnGluLysAspLysCysValGluAsn 381
Db 865 AGAGAAGATGTAATGTTTAAATTAACAAACAGAGGTGATGATGTTGTAATAT 924
Qy 382 ProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGlu 401
Db 925 CCAATCCTACTTGAACGAAATAATGGTGGATGTTGATGCGAGATGCCAATGTACCGAA 984
Qy 402 GluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLysProAspSerTyr 421
Db 985 GAAGATTCAGTACACGGAAGAAATACATGTGAATGTACTAAACCTGATTCTTAT 1044
Qy 422 ProLeuPheAspGlyIlePheCysSerSer 431
|||||
```

```
Db 1045 CCACCTTTTCGATGGTATTCTTCTGCAGTTCC 1074
RESULT 7
US-10-082-018-1
; Sequence 1, Application US/10082018
; Patent No. US20020144299A1
; GENERAL INFORMATION:
; APPLICANT: CHEN, LI How
; APPLICANT: MEADE, Henry
; TITLE OF INVENTION: NOVEL MODIFIED MSP-1 NUCLEIC ACID SEQUENCES AND
; TITLE OF INVENTION: METHODS FOR INCREASING mRNA LEVELS AND PROTEIN
; TITLE OF INVENTION: EXPRESSIONS IN CELL SYSTEMS
; FILE REFERENCE: 107.637.121A
; CURRENT APPLICATION NUMBER: US/10/082.018
; PRIOR FILING DATE: 2002-02-20
; PRIOR APPLICATION NUMBER: US/09/175,684
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1065
; TYPE: DNA
; ORGANISM: preferably, a bacterium, virus, or parasite
US-10-082-018-1
Alignment Scores:
Pred. No.: 1,5e-98 Length: 1065
Score: 1161.50 Matches: 218
Percent Similarity: 73.71% Conservative: 54
Best Local Similarity: 59.08% Mismatches: 76
Query Match: 51.05% Indels: 21
DB: 12 Gaps: 2
US-10-057-531a-2 (1-431) x US-10-082-018-1 (1-1065)
Qy 63 MetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeu 82
Db 19 ATCGATACATCTCTGTCCTCAAGATCGAAGACGAGTACGAGGTGCTGTCTCAAGCGCTG 78
Qy 83 AlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIlePheThrPheAsnLeu 102
Db 79 GCAGGGTCTACCGGAGCCTCAAGACGAGTGGAGAACACAGGTGATGACCTTCAACGTG 138
Qy 103 AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeu 122
Db 139 AACGTGAAGGATATCTCTGAACAGCGGTTCACAAAGCGGAGAACTTCAAGAAGCTGCTG 198
Qy 123 GluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSer 142
Db 199 GAGACGATCTGATCCCTTACAAAGGATCTGACACGAGTACGCAACACTACGTGGTCAAGGATCCC 258
Qy 143 PheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLys 162
Db 259 TACAAGTCTCTGACAGGAGAGAGAGATAGTTCCTGACGAGTACACTACATCATCAAG 318
Qy 163 GluSerValGluAsnAspIleLysPheAlaGlnGluLysIleSerTyrTyrGluLysVal 182
Db 319 GATAGCATTTGATACCGGATATCAACTTCGCCAACGATGCTCTGGGATACTACAAGATCTCTG 378
Qy 183 LeuAlaLysTyrLysAspAspLeuGluSerIleLysValIleLysGluLysGlu 202
Db 379 TCCGAGAGTACAAAGCGGATCTGGATTCAATCAAGAAGTACATC----- 423
Qy 203 LysPheProSerSerProProThrProProSerProAlaLysThrAspGluGlnLys 222
Db 424 -----AAGCAACAACGGA 438
Qy 223 LysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuVal 242
Db 439 GAGAACGAGAGATACCTGCTCTCTCAACAACATCGAGACCTGTACAGACCGTCAAC 498
Qy 243 AsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGlu 262
|||||
```

Db	19	ATTGATACACTACTTTCTAAAAATTGAAATGAATATGAGGTTTATATATTTAAACCTTTA	78
Qy	83	AlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeu	102
Db	79	GCAGTCTTTATAGAAGTTTAAAAAACAATTAGAAAAATACGTTTATGCATTTAATGTT	138
Qy	103	AsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeu	122
Db	139	AATGTTAAGGATATTTAAATTCACGATTTAATAAACGTTGAAATTTCAAAATGTTTAA	198
Qy	123	GluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSer	142
Db	199	GAATCAGATTTAATTCATATAAAGATTTAAACATCAAGTAATATGTTGTCAAGATCCA	258
Qy	143	PheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysLysSerTyrLysTyrIleLys	162
Db	259	TATAAATTTCTTAATAAAGAAAAAAGAGATAAATCTTAAAGCAGTTATATATATTAA	318
Qy	163	GluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLysVal	188
Db	319	GATTCAATAGATACGGATATAAATTTTGCAAAATGATGTTCTTGGAATATATATAAAT	378
Qy	183	LeuAlaLysTyrLysAspLeuGluSerIleLysLysValIleLysGluLysGlu	202
Db	379	TCCGAAAAATATAAATCAGATTTAGATTCAAATTAATAAATATATC-----	423
Qy	203	LysPheProSerProThrThrProProSerProAlaLysThrAspGluGlnLys	222
Db	424	-----AACGACAAACAGGT	438
Qy	223	LysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuVal	242
Db	439	GAATGAGAANAATACCTTCCTCTTTTAAACAATATTGAGACCTTATATAAAACAGTTA	498
Qy	243	AsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGlu	262
Db	499	GATAAATGATTTATTTGTAATTCATTTAGAAAGCAAAAGTTCTAAATTTATACATATG	558
Qy	263	LysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAspLys	282
Db	559	AAATCAACGTTAGAGTTAAATAAAGAAGCTTAATTAATCAAAACAATTCACACAAA	618
Qy	283	IleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAsp	302
Db	619	TTGCGAGATTTTAAAAAAAATAACAATTTGCTTGGAAATTCGCTGATTTATCAACAG	678
Qy	303	ThrLysLysAspMetGluLysLeuLeuSerThrGlyLeuVal--ClnAsnPhePro	321
Db	679	AACCATATAACTATTATGACAAAGTTCCCTTAGTACAGGTATGGTTTTGAAATCTT	738
Qy	322	AsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSerGln	341
Db	739	AAACCGTTTATCTAATTTACTTTGATGGAACTTCGCAAGGTATGTTAAACATTTTCA	798
Qy	342	HisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGlu	361
Db	799	CACCAATCGCTAAAAAAACAATGTCCACAAAATTCGTGGATGTTTCAGACATTTAGAT	858
Qy	362	ArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsn	381
Db	859	AGAGAAGAATGTAATGTTTATTAAATTTACAACAAGAAGGTGATAAATGTGTGAAAT	918
Qy	382	ProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGlu	401
Db	919	CCAATCCTACTCTACGAAAATATATGTTGGATGTGATGCAGATGCCAATATGATACCA	978
Qy	402	GluAspSerGlySerArgLysLysIleThrCysGluCysThrLysProAspSerTyr	421
Db	979	GAAGATTACAGGTACCAACGAGAAAGAAATCAACATGTGAATGTACTAAACCTGATCT	1038
Qy	422	ProLeuPheAspGlyIlePheCysSer	430
Db	1039	CCACTTTTCGATGCTATTTCTCGAGT	1065


```
QY 184 AlaLysTyrLysAspLeuGluSerIleLysLysValIleLysGluLys----- 201
Db 1147 CCGAGCTAT-----GGCTAGAGTCCATTTCAGAGGCTAATTGCCAGGTCATCTATTCT 1200
QY 202 ---GluLysPheProSer-----SerProThr-Th 211
Db 1201 CTAAATAATTCCTCAACAGAGAAACATTGTCAATCTCCTGGAGGCCAGTTGACTTAC 1260
QY 211 rProProSerProAla-----LysThrAspGluGlnLysLysGluSerLysPheLeuPr 229
Db 1261 CCCAGCCACTGCTGCTTTTGAACACTTGCACAAACAAAGACAGAAATTTTCACATTC 1320
QY 229 oPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeu 241
Db 1321 ATTTCTG-----AAAACCTTTCCAAACAAATGTG 1348

RESULT 19
US-10-046-583A-6
; Sequence 6, Application US/10046583A
; Patent No. US20020168743A1
; GENERAL INFORMATION:
; APPLICANT: Rice, John
; APPLICANT: Kloti, Andreas
; APPLICANT: Crawford, John
; APPLICANT: Lanning, Beth
; APPLICANT: Stewart, Sandy
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: IDENTIFICATION OF MODULATORS OF DEOXYXYLULOSE 5-PHOSPHATE
; FILE REFERENCE: 2037 US Divisional
; CURRENT FILING DATE: 2001-10-26
; PRIOR FILING DATE: 2000-07-27
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Nucleotides 1-495 encode the thioredoxin sequence
; OTHER INFORMATION: found in the vector PET32 supplied by No. US20020168743A1agen.
; OTHER INFORMATION: Nucleotides 496-2472 represent the tDXPS cDNA
; OTHER INFORMATION: sequence from Arabidopsis thaliana.
US-10-046-583A-6

Alignment Scores:
Pred. No.: 2,88e-16 Length: 2472
Score: 278.50 Matches: 76
Percent Similarity: 49.75% Conservative: 25
Best Local Similarity: 37.44% Mismatches: 49
Query Match: 12.24% Indels: 53
DB: 9 Gaps: 7

US-10-057-531A-2 (1-431) x US-10-046-583A-6 (1-2472)
QY 1 MethHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
Db 346 ATGCACCATCATCATCATCTTCTGCTCGGTGGCCAGCGGTCTCTGTATGAAGAA 405
QY 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40
Db 406 ACCGCTGCTGCTAAATTCCAGCCAGCACATGGACAGCCAGATCTGGTACCGACGAC 465
QY 41 AspAspLysAlaMetAlaAspIleGlySerIleGluGlyThrMetAlaIleSer 60
Db 466 GACGACAGGCCATGGCTGATATCGGATCCGCTTCA----- 501
QY 61 ValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLys 80
Db 501 ----- 501
```

```
QY 81 ProLeuAla-----GlyValTyrArgSerLeuLysGlnIleGluLysAsnIlePhe 98
Db 502 ---CTTGCAGAGAGGGTGAATATTATTCAC---AACAGACCACCACTCCATTACTTGAC 555
QY 99 ThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPhe 118
Db 556 ACTATTAACTACCAATCCAC---ATGAAATAATCTTTCTGTCAAGGAAGTGAACAACATT 612
QY 119 LeuAspValLeuGluSerAspLeuMetGlnPheLys-----HisIle 132
Db 613 TCTGATGACGTGAGATCAGACGTGATC---TTTAATGTCTGCAAAACCCGGTGGACATTG 669
QY 133 SerSerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsn 152
Db 670 GGGTCAAGTCCTGGTGTGGGAG----- 693
QY 153 ThrLeuLeuLysSerTyrTyrIleLysGluSerValGluAsnAspIleLysPheAla 172
Db 694 ---CTTACTGTGGCTCTTCAATATCAATCTCCACAGACAAGATCTTTGGGAT 750
QY 173 GlnGluGlyIleSerTyrGluLysValLeuAlaLysTyrLysAspLeuGluSer 192
Db 751 GTTGGTCATCAGTCTTATCTCTCATAGATTCTTACTGGGAGAGAGGAAGATGCCTACA 810
QY 193 IleLysLys 195
Db 811 ATGAGGCAA 819

RESULT 20
US-10-012-896-1010
; Sequence 1010, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darriack
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Wantanabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012,896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1010
; LENGTH: 1065
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-012-896-1010

Alignment Scores:
Pred. No.: 2,71e-15 Length: 1065
Score: 262.50 Matches: 60
```

Percent Similarity: 57.38% Conservative: 10
Best Local Similarity: 49.18% Mismatches: 19
Query Match: 11.54% Indels: 33
DB: Gaps: 5

US-10-057-531A-2 (1-431) x US-10-012-896-1010 (1-1065)

```
Qy 1 MethHisHisHisHisHisHisSerGlyLeuValProArgGlySerGlyMetLysGlu 20
Db 346 ATGCACCATCATCATCATCTTCCTGGTCTGCTGCCACGCGTCTCTGGTATGAAGAA 405
Qy 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 40
Db 406 ACCGCTGCTGTAATTCGAAGCGGCACATGACAGCCAGATCTGGGTACCGAGC 465
Qy 41 AspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMetAlaIle--- 59
Db 466 GACACACAGGCCATGGACCGGCTG-----GTGACGCGATTCGGCACCTCGACGAGTCTAT 519
Qy 59 ----- 59
Db 520 TTGGCCAGTGGCAGCTTTCCCTGTGGTGCCTGCCATGCTGCTCCACAGGTG 579
Qy 60 ---SerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyAspValIleTyr 78
Db 580 GCGGTGTGACAGCTTCAGCGCGCTCACCAGGTTCT-----ACCTTCTCAGCCCTCAG 633
Qy 79 LeuLysPro-----LeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsn 96
Db 634 ATCTGCGCTACACACTGGCTCCCTCTACACCGG-----GAGACGAG 678
Qy 97 IlePhe 98
Db 679 GTGTTC 684
```

RESULT 21

US-10-098-514-15'
; Sequence 15, Application US/10098514
; Publication No. US20020194648A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra P
; APPLICANT: Christopher, David A
; APPLICANT: Vine, Benjamin
; APPLICANT: Su, Wei-Wen
; APPLICANT: Bugos, Robert
; TITLE OF INVENTION: PLASMODIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PRODUCE
; TITLE OF INVENTION: TRANSGENIC PLANTS
; FILE REFERENCE: A-71339/RFT/TAL/NBC
; CURRENT APPLICATION NUMBER: US/10/098,514
; CURRENT FILING DATE: 2002-08-06
; PRIOR APPLICATION NUMBER: US 09/500,376
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: US 60/274,599
; PRIOR FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 150
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
US-10-098-514-15

Alignment Scores:
Pred. No.: 1e-07 Length: 150
Score: 168.00 Matches: 48
Percent Similarity: 24.12% Conservative: 0
Best Local Similarity: 24.12% Mismatches: 1
Query Match: 7.38% Indels: 150
DB: Gaps: 2

US-10-057-531A-2 (1-431) x US-10-098-514-15 (1-150)

```
Qy 167 AsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLysValLeuAlaLysTyr 186
Db 3 ANTGATATTAAATTTGCACAGAGGTATAAGTTATTATGAAAGGTT----- 50
Qy 187 LysAspAspLeuGluSerIleLysLysValIleLysGluLysGluLysPheProSer 206
Db 50 ----- 50
Qy 207 SerProProThrThrProProSerProAlaLysThrAspGluGlnLysLysGluSerLys 226
Db 50 ----- 50
Qy 227 PheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuValAsnLysIleAsp 246
Db 51 -----AACATTGAGACCTTATACATACTAGTTAATAAATTTGAC 92
Qy 247 AspTyrIleLeuLeuAsnLeuLysAlaLysIleAsnAspCysAsnValGluLysAspGluAla 266
Db 93 GATTACTTT----- 101
Qy 267 HisValLysIleThrLysLeuSerAspLeuLysAlaIleAspLysIleAspLeuPhe 286
Db 101 ----- 101
Qy 287 LysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAspThrLysLysAsp 306
Db 101 ----- 101
Qy 307 MetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThrIleLeuSer 326
Db 101 ----- 101
Qy 327 LysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSerGlnHisGlnCysValLys 346
Db 101 ----- 101
Qy 347 LysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGluCys 365
Db 102 -----CCAGAAATTCGTGGATGTTTCAGACATTAGATGAAGAGAAGATGT 149
```

RESULT 22

US-09-815-242-4196
; Sequence 4196, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16


```
Db 1108 TTCTTTAAACAGGTATCCCATCGAAGAAATTCCTACTCAATGCACCTT-----ATC 1158
QY 127 MetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSerPheLysLeuLeu 146
Db 1159 AGACGCTTGCACACTGCTGTGTTAAATA-----GATATCTGT 1197
QY 147 AsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLysGluSerValGlu 166
Db 1198 AATTCAGAGACCAAAAGGATTACTCAAG-----AAGGTTAA 1236
QY 167 AsnAspIleLysPheAlaGlnGluGlyIleSerTyr----- 179
Db 1237 AATGCCATTTCAGCTACAGCAGAGAGATTCTCGGTATATATGCAATGATGCTCATGAA 1296
QY 180 -----GluLysValLeuAlaLysTyrLysAspAspLeuSerIleLysValIle 197
Db 1297 TTTTAAAGTCAGTGTGTGGACAGCTGAAGAAGATATGCAAAATTAATTAACACTGG 1356
QY 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
Db 1357 AAGACTGAACCTGTTTCTGTGAGAGAAATTCACCATATTTTCAGCTACACAGCATAC 1416
QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGlu---Thr 236
Db 1417 ACT-----TGCCCTGTTTATTACTAATTTGGAGTTTGAG 1449
QY 237 LeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIle 256
Db 1450 GTTCAGCACTCCATCATTTGAAGCATGTGGAGATTTCCCAAAAGACAGACTTT 1509
QY 257 AsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeu 276
Db 1510 AATGACCTCTCTATTGACCTTCTCTGTAGGAAAAAACCACTCCCTCTCT----- 1557
QY 277 LysAlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGlu---AlaIle 295
Db 1558 CGTTCAATTCAAGATTCTCTTGATCTTCTTTAGGCCCGAAGAACTGGAGTATCTTGT 1617
QY 296 LysLys-----LeuIleAsnAspAspThrLysLysAspMetLeuGly 309
Db 1618 GAGAGTGTGTGGAGTGTCTCTGTCTCAGGCAC-----AAATTTAACAGCTTCTCT 1671
QY 310 LysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThrIleIleSerLysLeuIle 329
Db 1672 AGGTCTCTATCTCCATTTGAACAGATATAGCTTCAATGTGGCTCTCTCG-----CTT 1725
QY 330 GluGlyLysPheGlnAspMetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCys 349
Db 1726 AACAATAAGATTGGGCAGCAAGTCATCTTCCAAGATACCTGACCCCTGTCTCATCTTC 1785
QY 350 ProGluAsnSer-----GlyCysPheArgHisLeuAspGluArgGlu 363
Db 1786 ACTGAAATACAAACACCACCTTTTACCCTTGGTGGAGTGCACATATGCAATGCTCTAGA 1845
QY 364 GluCysLysCysLeuLeuAsnTyrLysGlnGluLysAspLysValGluAsnProAsn 383
Db 1846 CCATTGAAAGCC-----TCTCAAAATGGTGAATCTCTGCATCCACAGCCCTTCT 1893
QY 384 ---ProThrCysAsnGluAsnAsnGlyGlyCysAspAlaThrCysThrGluGlu 402
Db 1894 ACACCTTCAAGAAATTCACCTTCAATCCAAAGAGCTCTTGGCTTTATGCTTGAATCA 1953
QY 403 AspSerGlySerArgLysLys 410
Db 1954 GACAGTGAGGATGAGCTAAACGT 1977
```

RESULT 25

```
US-09-974-300-2660
; Sequence 2660, Application us/09974300
; Patent No. US2002014671A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
```

```
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: Expression
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2660
; LENGTH: 3057
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2660

Alignment Scores:
Pred. No.: 0.00492 Length: 3057
Score: 137.00 Matches: 109
Percent Similarity: 36.97% Conservative: 57
Best Local Similarity: 24.28% Mismatches: 163
Query Match: 6.02% Indels: 120
DB: 10 Gaps: 24

US-10-057-531A-2 (1-431) x US-09-974-300-2660 (1-3057)
QY 8 SerSerGlyLeuValProArgGlySerGlyMetLysGluThrAlaAlaLysPheGlu 27
Db 1303 TCCTCAGGAATGAGCGGACAAATTCAGGAGATAAAGACCATCTGCGGAGAGACCCGAAT 1362
QY 28 ArgGlnHisMetAspSerProAspLeuGlyThrAsp---AspAspAspLysAlaMetAla 46
Db 1363 CAAGGACAAGAACCTTCCCTCGGAACCGGAAACGGTCAGCAAGCGCGCGCAACT 1422
QY 47 AspIleGlySerIleGluGlyArgGlyThrMetAlaIleSerValThrMetAspAsnIle 66
Db 1423 CCTTCAGATGAGATAAATGATCAAGGGACG----- 1452
QY 67 LeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeuAlaGlyValTyr 86
Db 1453 -----GGTGAAGATGATCAA-----CAGCTGAACCT----- 1479
QY 87 ArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeuAsnLeuAsnAsp 106
Db 1480 -----GGAAACGAACACCGGAAGATGAG-----AATGCTAAGCTTAAAGAA 1521
QY 107 IleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeuGluSerAspLeu 126
Db 1522 CAATTGAACGAAGCGCGCAAGG-----CTTGAATAATTTGGAAGAAGATCTT 1569
QY 127 -----MetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSerPheLys 144
Db 1570 CGTGCCTAAACAGACACTCATACGAGAGCTGGAACAAACAGATTGAAATTAATAAG 1629
QY 145 LeuLeuAsnSerGluGlnLysAsn-----ThrLeuLeuLysSerTyrLysTyrIleLys 162
Db 1630 AAATTAGTACTCTGAAGAAACACATTAAAGCATTAAGCATGAATCAAAAGATTGAAG 1689
QY 163 GluSerValGluAsn-----AspIleLysPheAla 172
Db 1690 GTGCATATTGAGATCTCAATAAATAATTTTGTATGATCTATGATGATTTAAAAAATTT 1749
QY 173 GlnGluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSer 192
Db 1750 GAAATGGCATATTAGAGTAT-----CTCCGAAAGATAGTAGTCGGAAGAGAAAC 1800
QY 193 IleLysLysValIleLysGluGluLysPheProSerSerProProThrThrPro 212
Db 1801 TTAGCAAGGCTTTTGAG----- 1818
QY 213 ProSerProAlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThr 232
Db 213 ProSerProAlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThr 232
```



```
Db 1720 AAATCCTTGCAGGCTGACACC-----ACCAACACTGAC----- 1752
Qy 225 SerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuValAsnLys 244
Db 1753 -----ACTGCCTTGACAACCTTTGGAGGAGGCCCTTGCAGAGAAA 1791
Qy 245 IleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGluLysAsp 264
Db 1792 ---GAGCGGCAATTTGAACGCTTAAGAGCAGAGGAGCAGAGATGAGCGAGAGAGCAA 1848
Qy 265 GluAlaHisValLysIleThrLeuSerAspLeuLysAlaIleAspAspLysIleAsp 284
Db 1849 GAGGAATGATACTACAAAAA-----GATCTTAAAGACTTTGAAGGAAAAAGTCAAGC 1902
Qy 285 LeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAspAspThrLys 304
Db 1903 CTGTTGCAAGCGGACCTTTCAGAGAAAGAGCGCTTCACTTTG-----GATCTG 1950
Qy 305 LysAspMetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThrIle 324
Db 1951 AAAGAGCATGCTTCTCTGCGCATCTCAGGACTGAAAGAGGACTCAGCGCTTAAGACA 2010
Qy 325 IleSerLysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSerGln----- 341
Db 2011 CTAGAGATTGCTTTGGAGCAGAGAGGAGGTGCTGAAATGGAATCAACAATTGAAA 2070
Qy 342 -----HisGlnCysVal-----LysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 2071 AAGGCACATGAGGAGCAGCATGGAAGCCAGAGCCAGTCCAGAGATGAGTGACCGCAATACAG 2130
Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGly----- 375
Db 2131 CACTTG---GAGAGAG-----ATCACCAGGTACAAAGATGATCTAGCAAG 2175
Qy 376 -----AspLysCysValGluAsnProAsnProThrCysAsnGluAsnAsn 390
Db 2176 GCCCAGCAGAGAAGTTGATCGACTCTAGAAATCTTGAAGGAGGTGGAATAATGAGAAGAA 2235
Qy 391 GlyGlyCysAspAlaAspAlaThrCysThrGluGluAspSerGlySerSerArgLys 409
Db 2236 -----GACAAGATAGAAGATAGCTGAGTTGGAAGACTCTCACCTCAAGGCAA 2283

RESULT 27
US-10-267-311-20
: Sequence 20, Application US/10267311
: Publication No. US20030050469A1
: GENERAL INFORMATION:
: APPLICANT: Siegel, Marvin
: APPLICANT: Chu, N. Randall
: TITLE OF INVENTION: INDUCTION OF A THI-LIKE RESPONSE IN VITRO
: FILE REFERENCE: 12071/002001
: CURRENT APPLICATION NUMBER: US/10/267,311
: PRIOR APPLICATION NUMBER: 2002-10-09
: PRIOR FILING DATE: 2000-07-10
: PRIOR APPLICATION NUMBER: US 60/143,757
: PRIOR FILING DATE: 1999-07-08
: NUMBER OF SEQ ID NOS: 55
: SOFTWARE: FastSEQ for Windows Version 4.0
: SEQ ID NO 20
: LENGTH: 2847
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: fusion sequence
: NAME/KEY: CDS
: LOCATION: (1)...(2844)
US-10-267-311-20
```

Alignment Scores:

```
Pred. No.: 0.0144 Length: 2847
Score: 131.50 Matches: 82
Percent Similarity: 35.12% Conservative: 62
Best Local Similarity: 20.00% Mismatches: 123
Query Match: 5.78% Indels: 143
DB: 9 Gaps: 16

US-10-057-531A-2 (1-431) x US-10-267-311-20 (1-2847)
Qy 2 HisHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGluThr 21
Db 13 CATCATCATCATCATCATCAGCAGCGCGCTGGTCCGCCGCGCAGCATATGCCAAGACA 72
Qy 22 AlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAspAsp 41
Db 73 ATTCCGTACGACGAGAGGCCCGTCGC-----GGCTCGAGCGGGGC 114
Qy 42 AspLysAlaMetAlaAspIleGlySerIle-----GluGlyArgGlyThrMet 57
Db 115 TTGAACGCCCTCGCCGATGCGGTAAAGGTGACATTGGGCCCGCCAGGGCGCAACGTCGTC 174
Qy 58 -----AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGlu 71
Db 175 CTGGAAGAAGAGTGGGTGGCCCGATCACCAACGATGCTGTCTCATCCCAAGGAG 234
Qy 72 AsnGluTyrAspValIleTyrLeuLys----- 80
Db 235 ATCGAGCTGGAGGATCCGTACGAGAGATGCGGCCGCGAGTGGTCAAGAGGTAGCCCAAG 294
Qy 80 ----- 80
Db 295 AAGACCGATGACGTGCGCGGTGACGGCACACACGACGCGCCGCTGCTGGCCGAGCGTTG 354
Qy 81 -----ProLeuAlaGlyValTyrArg 87
Db 355 GTTCGCGAGGCGCTCGCAACGTCGCGCGCGCGCCGCAACCGCTC---GGTCTCAACGC 411
Qy 88 SerLeuLysLysGlnIleGluLysAsnIlePheThrPhe----- 100
Db 412 GGCATCGAAGAGCGCGTGAGAGGTCAACGAGACCTGCTCAAGGGCGCCCAAGAGGTC 471
Qy 101 -----AsnLeuAsnLeuAsn 105
Db 472 GAGACCAAGAGCAGATTTGCGGCCACCGACGATTTCCGGCGGTGACAGTCCATCGGT 531
Qy 106 AspIleLeuAsnSerArgLeuLysLys-----ArgLysTyrPheLeuAspValLeuGlu 123
Db 532 GACCTGATCGCGGAGCGCATGACAAGGTGGGCAACGAGGGCGTCTACCGTTCGAGGAG 591
Qy 124 SerAspLeu-----MetGlnPheLys-----His 131
Db 592 TCCAACACCTTTGGCTGCGAGCTCGAGCTACCGAGGGTATGCGGTTCCACAGGGCTAC 651
Qy 132 IleSerSer-----AsnGluTyrIleIleGluAspSerPhe 143
Db 652 ATCTCGGGGTACTTCGTGACCGACCGCGGAGCGTTCAGAGAGCGGCTCTGAGAGCCCTAC 711
Qy 144 LysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLysGlu 163
Db 712 ATCTCTGTGTCACTCAAGGTGTCCACTGTC----- 744
Qy 164 SerValGluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLysValLeu 183
Db 745 -----AAGGATCTGCTCCCGCTCTCGAAGAGGTGTCATC 777
Qy 184 AlaLysTyrLys-----AspAspLeuGluSer-----IleLys 194
Db 778 GGAGCCGGTAAAGCGCGCTGCTGATCATCGCCGAGGACGTCGAGGGCGGAGCGGTGTCCACC 837
Qy 195 LysValIleLysGluGluLysPheProSerSerProThrThrProProSer 214
Db 838 CTGCTGTCAACAAGATCCCGCGGCACTTCAAGTCG-----GTGGCGGTCAAGGCT 888
```



```
; LENGTH: 2341
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (966)...(2291)
US-09-881-752A-147

Alignment Scores:
Pred. No.: 0.0187 Length: 2341
Score: 129.00 Matches: 89
Percent Similarity: 41.58% Conservative: 64
Best Local Similarity: 24.18% Mismatches: 125
Query Match: 5.67% Indels: 91
DB: 10 Gaps: 19

US-10-057-531A-2 (1-431) x US-09-881-752A-147 (1-2341)
QY 85 ValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeuAsnLeu 104
Db 562 GTAAAGAGAAGTTTAAAGAAAAATCCTTAAGGAATTTGAAAAACGATAAATAC-AATCAA 620
QY 105 AsnAspIleLeuAsn---SerArgLeuLysLysArg----- 115
Db 621 AGCGAATAGTAGGGTTAGAAAAATTTAAAGGAAAAATTCAGATTCGTTTGGTGAAAAAC 680
QY 116 LysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPhe----- 129
Db 681 CAAACAGAATTGGCACTATTGGAATCCAAATTTAACAGATTTTCATTTTATTGAAATCAT 740
QY 130 -----LysHisIleSerSerAsnGluTyrIleIleGluAspSerPheLys 144
Db 741 TCTATTGGGAACAAAAATTTAGGAGCTGGTGATCGACCATTCGACATTTAATAAAAA 800
QY 145 LeuLeuAsnSerGluGlnLysAsnThrLeuLysSerTyrLysTyrIleLysGluSer 164
Db 801 AGATTAAAGCAATGAA-----GATCGGTAGCTCAGGTAGAGATATATATAAAA----- 848
QY 165 ValGluAsnAspIle---LysPheAlaGlnGluGlyIleSerTyrTyrGluLysValLeu 183
Db 849 ---GATAATAGTATATATGCGCTTCTGTCAAAA-----GAAACCATTTACC 890
QY 184 AlaLysTyrLysAspAspLeuGlu-----SerIleLysLysValIleLys 198
Db 891 GAAGAATTTAAAAACAACATAGAACTTATTGATACAAAGTTATCAAGAATCTATTGAA 950
QY 199 GluGluLysGluLysPheProSerProProThrThrProProSerProAlaLysThr 218
Db 951 ACGATCAAGGAAAGATGGAAGACTACGACGAGAACCGCTGGGACCTGGAGCCGACTT 1010
QY 219 Asp-----GluGlnLysLysGluSerLysPheLeuProPheLeuThrAsn 233
Db 1011 GATAAGATTGTTGAAACAGACAGAGAAGATCAACAAACTAAATTTGAC---ACAGAAAAAT 1067
QY 234 IleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeu--- 252
Db 1068 TTGAAAAATAATTATTGAAACTTTTGAGAAGTAAATCAATGGGAATCAGCAAAAGATGCTT 1127
QY 253 ---LysAlaLysIleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThr 271
Db 1128 GATAAAGTAAAGAAATGACGAGAAATTTTAAGCTTGATAGCACTAAAACGAGATAGAC 1187
QY 272 LysLeuSerAspLeu---LysAlaIleAspAspLysIle-----AspLeuPhe 286
Db 1188 GCAATTAAGATTGTGATTTAAAAAGGCTAATGAGCAAAATGACCAATTAATGAGATGATA 1247
QY 287 LysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAspAspThrLysLysAsp 306
Db 1248 AAG-----GATATTGAAAAACAGAAAAAGAGTTGTAAGGAACAACTTGGAAATTT 1298
QY 307 MetLeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThrIleIleSer 326
Db 1299 CTAGTCAATGATTTTAAAGT----- 1319

; Sequence 40, Application US/10098841
; Publication NO. US20020197679A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Xu, Chongjun
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-Hong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
; FILE REFERENCE: 784CIP2
; CURRENT APPLICATION NUMBER: US/10/098,841
; PRIOR FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 331
; SOFTWARE: pc_FL-genes Version 1.0
; SEQ ID NO. 40
; LENGTH: 6386
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (150)..(5492)
US-10-098-841-40

Alignment Scores:
Pred. No.: 0.0755 Length: 6386
Score: 129.00 Matches: 81
Percent Similarity: 36.02% Conservative: 53
Best Local Similarity: 21.77% Mismatches: 114
Query Match: 5.67% Indels: 124
DB: 9 Gaps: 16

US-10-057-531A-2 (1-431) x US-10-098-841-40 (1-6386)
```

```
Qy 71 GluAsnGluTyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLys 90
|||
Db 2253 GAAATGAATCAGAT-----TCATTGATT 2276

Qy 91 LysGlnIleGlu-----LysAsnIlePheThrPheAsnLeuAsnAspIle 107
|||
Db 2277 CAAGAGCTTGACATCTAATAAGAAATAATATACAGAAATCAAGAAATAAGAAATG 2336

Qy 108 LeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeuGluSerAspLeuMet 127
|||
Db 2337 ATAAT-----ATAATTGATCAAAAGAGAGATCTATCAAC 2372

Qy 128 GlnPheLysHisLeuSerSerAsnGluTyrIleIleGluAspSerPheLysLeuAsn 147
|||
Db 2373 GAATTCAGAACCTAAAGTCTCAT-----ATGGAACACATTTAAATGCAATGAC 2423

Qy 148 SerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIle-----LysGluSerValGlu 166
|||
Db 2424 AAGCTGATACATCTCTTTAATAATAAACAATAAATGATTTGTAATGAACAGTTGAA 2483

Qy 167 AsnAspIleLysPheAlaGlnGluGlyLysSerTyrTyrGluLysValLeuAlaLysTyr 186
|||
Db 2484 -----GTACCTAAGCAGACG 2498

Qy 187 LysAspAspLeuGluSerIleLysLysValIleLysGlu-----GluLysGluLys 203
|||
Db 2499 AATCTAAATCTCTTCAGAAAGAAAGAGAGTAATGAATGAATCACTTCAGCAAGATGAA 2558

Qy 204 PheProSerSerProProThrThrProProSerProAlaLysThrAspGluGlnLysLys 223
|||
Db 2559 CCACCACCAAGAAAGGCTCTATCATGTTAGTTTCAGCTATCACTGAAGACCAAAAGAAA 2618

Qy 224 GluSerLysPheLeuProPheLeuThrAsnIleGlu----- 235
|||
Db 2619 AGTGAAGAAGTCCACCGCAACATTCGAGAAATTCGAACATACATCAGAGTTTACAGAAAT 2678

Qy 236 -----ThrLeuTyrAsnAsnLeuValAsnLysIleAsp 246
|||
Db 2679 AATGAAGGACTGAGAGCATTTTACTCACTATTGAGAATGAACCTTAAATAAGAAAGGAA 2738

Qy 247 AspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGluLysAspGluAla 266
|||
Db 2739 GAA-----AAGCAGAATTAAT-----AAACAGATTGTT 2768

Qy 267 HisValLys-----ileThrLysLeuSerAspLeu 276
|||
Db 2769 CATTTTCAGCAGGAACCTTCTCTTTCTGMAAAAGAAATTTAACTTTAAGTAAGAGGTC 2828

Qy 277 LysAlaIleAspAspLysIleAspLeu-----PheLys 287
|||
Db 2829 CAACAAATTCAGTCAAAATTAATGATATTGCAATTTGCTGAATTTACATGTGCAGAAAGTAA 2888

Qy 288 AsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAsp----- 301
|||
Db 2889 AATCAAGAACAGGAGGAAAGATCATGAAATTTGCAATGAGATGAGAACTGCTACAGA 2948

Qy 302 -----AspThrLysLysAspMetLeuGly 309
|||
Db 2949 AGCATTACAATAATGTTTCAACAATAAATTAATGACACAGAAATAAGAGCAACTAGCT 3008

Qy 310 LysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThrIleIleSerLysLeuIle 329
|||
Db 3009 ACTCTTGATTCAGTTCTCAGATTTCACATAGATTGCTCAATCTCAGGGATCTGCA 3068

Qy 330 GluGlyLysPheGlnAspMetLeuAsnIleSerGln-----HisGln 343
|||
Db 3069 AATGGTTCTGAGGAGGATAATTTGCCAAATACACAGTTAGACCTTTTAGTAAATGATTAT 3128

Qy 344 CysValLysLysGlnCys-----ProGluAsnSerGlyCysPhe 356
|||
Db 3129 TTGTAAGTAAGCAAGTTAAGAAATATCGAATTCAGAACCCCAATAGGGAATAATCTTTC 3188
```

```
Qy 357 ArgHisLeuAspGluArg-----GluGluCysLys 366
|||
Db 3189 CACTCTAGTATTGAAGCTATTTGGGAAGAAATGTAA 3224

RESULT 31
US-10-046-935-2238
; Sequence 2238, Application US/10046935
; Patent No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2238
; LENGTH: 1176
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-046-935-2238

Alignment Scores:
Pred. No.: 0.0089 Length: 1176
Score: 128.00 Matches: 82
Percent Similarity: 33.41% Conservative: 64
Best Local Similarity: 18.76% Mismatches: 171
Query Match: 5.63% Indels: 120
DB: 9 Gaps: 16

US-10-057-531A-2 (1-431) x US-10-046-935-2238 (1-1176)
Qy 2 HisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGluThr 21
|||
Db 13 CATCATCATCATCATCACAGCGCGCTGTGCGCGCGCGCATATG----- 63

Qy 22 AlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 41
|||
Db 64 -----GACGCTCGCGCTGCGCGAGAAAGATCTCAGAGTAAAGAAAGAAC 108

Qy 42 AspLysAlaMetAlaAspIleGlySerIleGluGlyThrMetAlaIleSerVal 61
|||
Db 109 TTAAGAAATTCAGATATGGAAGTTGATT-----TCCATGGAACCTCGTCA 156

Qy 62 ThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysPro 81
|||
Db 157 TCCTCTGATGACACTGTGACAGCTTTGCTTCTCAT----- 192

Qy 82 LeuAlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIlePheThrPheAsn 101
|||
Db 192 ----- 192

Qy 102 LeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspVal 121
|||
Db 193 -----AATTTGCAACACAGCGCTGCAG-----TCAGTT 222

Qy 122 LeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAsp 141
|||
Db 223 CGGAAGGCTGTAGAGCCCGCAGCGACTGCGAGGCACTCTGACCTCTCAGGGTGGCGATG 282

Qy 142 SerPheLysLeuLeuAsnSerGlu-----GlnLysAsnThrLeuLeuLysSerTyrLysTyr 160
|||
Db 283 AAGTTTCCAGCGCGGATACAGGAGCAACCAACCAAGCAAGAGTCCCGCACGCC 342

Qy 161 IleLysGluSerVal-----GluAsnAspIleLysPheAlaGlnGluGlyIleSerTyr 178
|||
Db 343 TCAGAGAAATCTGTGACTGATTCACATCCGATTCAGAAAGTGAAGTGAATTTT 402
```

```
QY 179 TyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysValIleLys 198
Db 403 TTGGAGAAAGGGCTTTAAATAAAGCAAAACAAACCAATGCTTGCAAAACTCATG--- 459
QY 199 GluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLysThr 218
Db 460 TCTGAATTAGAAAGCTTCCCTGGCTCGTTCGCGTGAAGACATCCCTCCCGAGCTCCGAC 519
QY 219 AspGluGlnLysLys----- 223
Db 520 TCACAATCAAGGAGAGCCGGAAGCGGTACATTCCCGGGTGTGCTTCCAGAGAAACCCCT 579
QY 224 GluSerLysPheLeuPro-----PheLeuThrAsnIleGluThr 236
Db 580 GAACGGAGAGCTCGCTCTTACCAGGTCAAGGTCCCGGATCCCGGTCCCTTGACGCT 639
QY 237 LeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIle 256
Db 640 CTACCCATGAGGAGGAGGAGGAGGAGTAAGTACATGTTGGTGAGAAAGAGGAAGACC 699
QY 257 AsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeu 276
Db 700 GTGGATGGCTACATGAATGAAGTAC-----CTGCCAGAGAC 738
QY 277 LysAlaIleAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLys 296
Db 739 CGTGGCTCCAGATCATCCGTGACCCCT-----CCGCATATAATTCCGCCAGTGGAA 789
QY 297 LysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeu 316
Db 790 GAAATTACAGAGGAGGAGTTGGAG----- 813
QY 317 ValGlnAsnPheProAsnThrIleIleSerLysLeuIleGluLysPheGlnAspMet 336
Db 814 -----ACGCTGCGAGCAATCTCGAGAGAGATATATAACCGTTCA 855
QY 337 LeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPhe 356
Db 856 CTGGGCTCTACTTGTCTCAATGCGTCAAGAGACTATTGATACCAAAACAAACTGC--- 912
QY 357 ArgHisLeuAspGluArgGluGluCys-----LysCys 367
Db 913 -----AGAAACCCAGACTGCTGGGGGTTTCGAGGCCAGTTCTGTGGCCCCCTGC 960
QY 368 LeuLeuAsn---TyrLysGlnGluGlyAspLysCysValGluAsnProAsnProThrCys 386
Db 961 CTTGAAACCGTTATGTGTGAAGAGTCCAGGATGCTCTGTGGATCCGAACTGGCATTCG 1020
QY 387 AsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGluGluAsp 403
Db 1021 CCGCCTTGTGAGGAATCTGCAACTGC---AGTTTCTGCGGCGAGAGAT 1068

RESULT 32
US-10-146-502-2238
; Sequence 2238, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C2
; CURRENT APPLICATION NUMBER: US/10/146,502
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2238
; LENGTH: 1176
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
US-10-146-502-2238
Alignment Scores:
Pred. No.: 0.00889
Score: 128.00
Percent Similarity: 33.41%
Best Local Similarity: 18.76%
Query Match: 5.63%
Indels: 120
Gaps: 16

US-10-057-531A-2 (1-431) x US-10-146-502-2238 (1-1176)
QY 2 HISHISHISHISHISerSerGlyLeuValProArgGlySerGlyMetLysGluThr 21
Db 13 CATCATCATCATCATCAGCAGCGGCTGTGTCGCGCGCGCATATG----- 63
QY 22 AlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThrAspAsp 41
Db 64 -----GACGCTCGCGCGCTGCGCAGAAAGATCTCAGAGTAAAGAAAC 108
QY 42 AspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMetAlaIleSerVal 61
Db 109 TTAAGAAATTCAGATATGTAAGTTGAT-----TCCATGGAAACCTCGTCA 156
QY 62 ThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysPro 81
Db 157 TCCTCTCATGACAGTGTGACAGCTTTGCTTCTGAT----- 192
QY 82 LeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsn 101
Db 192 ----- 192
QY 102 LeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspVal 121
Db 193 -----AATTTTGCAACACGAGGCTGCAG-----TCAGTT 222
QY 122 LeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAsp 141
Db 223 CGGGAAGGCTGTAGGACCCGCGCAGCGCTGCGAGGCACTCTGAGACCTCTCAGGTTGGCGATG 282
QY 142 SerPheLysLeuLeuAsnSerGlu---GlnLysAsnThrLeuLeuLysSerTyrLysTyr 160
Db 283 AGTTTCCAGCGGAGTACCGGGGAGCAACCAACAAAGACAGATCCCGCCAGCC 342
QY 161 IleLysGluSerVal-----GluAsnAspIleLysPheAlaGlnGluGlyIleSerTyr 178
Db 343 TCAGAGAATTCGTGTGACTGATTCACACTCCGATTCAAGAGATGAAGTGAATTT 402
QY 179 TyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValIleLys 198
Db 403 TTGGAGAAAGGGCTTTAAATAAAGCAAAACAAACCAATGCTTGCAAAACTCATG--- 459
QY 199 GluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLysThr 218
Db 460 TCTGAATTAGAAAGCTTCCCTGGCTCGTTCGCGTGAAGACATCCCTCCCGAGCTCCGAC 519
QY 219 AspGluGlnLysLys----- 223
Db 520 TCACAATCAAGGAGAGCCGGAAGCGGTACATTCCCGGGTGTGCTTCCAGAGAAACCCCT 579
QY 224 GluSerLysPheLeuPro-----PheLeuThrAsnIleGluThr 236
Db 580 GAACGGAGAGCTCGCTCTTACCAGGTCAAGGTCCCGGATCCCTCGGTCCTCTTGACGCT 639
QY 237 LeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIle 256
Db 640 CTACCCATGAGGAGGAGGAGGAGGAGTAAGTACATGTTGGTGAGAAAGAGGAAGACC 699
QY 257 AsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLysSerAspLeu 276
Db 700 GTGGATGGCTACATGAATGAAGTAC-----CTGCCAGAGAC 738
QY 277 LysAlaIleAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLys 296
```


Db 739 CCGTCCCTCAGATCCGCTGACCCTT-----CCGATATATATTCGCCCGCCAGTGGA 789
Qy 297 LysLeuLeuAsnAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeu 316
Db 790 GAAATTACAGAGGAGGAGTTGGAG-----
Qy 317 ValGlnAsnProAsnThrIleLeuSerLysLeuLeuGluGlyLysPheGlnAspMet 336
Db 814 -----AACCTGTCGAGCAATTCGAGAGAGATATATATACCGCTTCA 855
Qy 337 LeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPhe 356
Db 856 CUGGCTCTACTGTCTCACTCAATGCGCTCAGAGACTATTGATACCAAAACAACTGC--- 912
Qy 357 ArgHisLeuAspGluArgGluGluCys-----LysCys 367
Db 913 -----AGAAACCCAGACTGCTGGCGCTTCGAGGCCAGCTTCTGTGCCCTGTC 960
Qy 368 LeuLeuAsn---TyrLysGlnGluGlyAspLysCysValGluAsnProAsnProThrCys 386
Db 961 CTTGAAACCGGTATGGTGAAGAGTCAAGGATGCTCTGCTGGATCCGAACTGGCATTCG 1020
Qy 387 AsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGluGluAsp 403
Db 1021 CCGCTTGTGAGGATCTGCAACTGC---AGTTTCTGCGCGCAGCGAGAT 1068

RESULT 33

US-09-964-824A-263

; Sequence 263, Application US/09964824A

; Patent No. US20020102531A1

; GENERAL INFORMATION:

; APPLICANT: Horrigan, Stephen

; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu

; FILE OF INVENTION: Sets

; CURRENT APPLICATION NUMBER: US/09/964,824A

; PRIOR FILING DATE: 2001-09-27

; PRIOR APPLICATION NUMBER: US/60/236,033

; PRIOR FILING DATE: 2000-09-28

; PRIOR APPLICATION NUMBER: US/60/236,032

; PRIOR FILING DATE: 2000-09-28

; PRIOR APPLICATION NUMBER: US/60/236,028

; NUMBER OF SEQ ID NOS: 583

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 263

; LENGTH: 4198

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-964-824A-263

Alignment Scores:

Pred. No.:	0.0645	Length:	4198
Score:	127.00	Matches:	70
Percent Similarity:	41.82%	Conservative:	68
Best Local Similarity:	21.21%	Mismatches:	112
Query Match:	5.58%	Indels:	80
DB:	10	Gaps:	17

US-10-057-531A-2 (1-431) x US-09-964-824A-263 (1-4198)

Qy 86 TyrArgSerLeuLys---LysGlnIleGluLysAsnIlePheThrPheAsnLeuAsnLeu 104
Db 1059 TACCAAGCAATGAAGAACTGACTCTTCGAGAGAGATGACAAATATTCTTTGACTGTG 1118
Qy 105 AsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeuGluSer 124
Db 1119 AATGATGTGAAGAACACT-----TACTCTCTCCCTAGGAAGA 1154
Qy 125 AspLeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIleGluAspSerPheLys 144
Db 1155 AAAGTCAGCGAAGATAAAA-----AGCAGAGAATTT-----CAATCTCTCTCTAAAA 1199

Qy 145 LeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLysGluSer 164
Db 1200 GGTCTAAATCCAAAGCATTAATGCTACTGATAAGAGAC-----ATAGTAAGACAACA 1253
Qy 165 -----ValGluAsnAspIleLysPheAlaGlnGluGlyLysSerTyrTyrGluLys 181
Db 1254 TTTAAATTTTTCAAAATACATG-----CAAGAGACTGTGACACAGCTCTTTCAAG 1304
Qy 182 ValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluGluLys 201
Db 1305 ACTGTATCAAGTCTATCAGAGACCTCGAAGACCACCAAGCAAAATTAATCAAAAAGTTAAT 1364
Qy 202 GluLysPheProSerSerProProThrThrProProSerProAlaLysThrAspGluGln 221
Db 1365 GAATCTGTGGTTTCAATA-----
Qy 222 LysLysGluSerLysPheLeu-----ProPheLeuThrAsnIleGlu 235
Db 1383 GCAGCCCGAGCAAAAGTTTGTGTCGCAAGAGAAATCGGCCACACTTTGACTGATATAGT 1442
Qy 236 ThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLys 255
Db 1443 GAACTAAGGAATCACATTTGTAATGTAAGCAAGAAATGACTCTTACATGTGTGAGAGCCT 1502
Qy 256 IleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAsp 275
Db 1503 ATTAAGAACTAGAAAGTAAAG-----CAGACTCATTTAGAA-----GGTGTCT 1544
Qy 276 LeuLysAlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIle 295
Db 1545 CTAGAACAGGAACTCAAGAGCAATTCGTGTAT-----TATGAATCCCTC 1589
Qy 286 LysLysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGly 315
Db 1590 AATAAACTCTTTCT-----AAATGAAGGAAGTACATGAGCAGCTTTTATCAACTGAA 1643
Qy 316 LeuValGlnAsnProAsnThrIleLeuSerLysLeuIleGluGlyLysPheGlnAsp 335
Db 1644 CAGGTATCAGACCAGGAAGATGCTCCAGCTGCTGAGTCAGTGTAGCAATAATGTCTACTGAG 1703
Qy 336 MetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCys 355
Db 1704 TACATGTCT---ACTTTACATGAAATATATAAGAAAGCAGAGATTGATGATGCTGCAATG 1760
Qy 356 PheArg-----HisLeuAspGlu-----
Db 1761 TTTGAAGATTTGCACATTTCAAGAAAGCAAGATTAACAATCTCACCGTCTCTTTGGAGATG 1820
Qy 362 -----ArgGluGluCys-----LysCysLeuLeuAsnTyr 371
Db 1821 GAGAAAGAGTCTCTCAGAGGTGAATGTGAAGACATGTTATCCAAATGCAGAAATGATTTT 1880

RESULT 34

US-10-044-090-102

; Sequence 102, Application US/10044090

; Patent No. US20020137081A1

; GENERAL INFORMATION:

; APPLICANT: Olga Bandman

; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION

; FILE REFERENCE: PR-0028 US

; CURRENT APPLICATION NUMBER: US/10/044,090

; CURRENT FILING DATE: 2002-01-09

; NUMBER OF SEQ ID NOS: 850

; SOFTWARE: PERL Program

; SEQ ID NO 102

; LENGTH: 4967

; TYPE: DNA

; ORGANISM: Homo sapiens

Qy	63	MetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIleTyrLeuLysProLeu	82
Db	1778	ATGCATATTTCCATAAGAATACTGAAGGGAAGAAAGATTGATCTT	1825
Qy	83	AlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeu	102
Db	1826	-----GAACCTGGGAAGAAGTAGAGCAACAATC---CAGTACAAACAGT	1867
Qy	103	AsnLeu-----AsnAspIle-----LeuAsnSerArgLeuLysLysArg	115
Db	1868	GACTAGAACAAAGGTAAATGAATTACAGGAGGACTAGAGGAGCTTTAAAGAAAG	1927
Qy	116	LysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsn	135
Db	1928	GATCAAAATGCCAAAACATAGAAAACATTATGGTTCAAATCAAGATTCCTCTCAAGAC	1987
Qy	136	GluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeu	155
		::: :::::	
Db	1988	AAAGAAGTATTGTCAAGTGAAGTCAAGTCTCTT---TATGAGGAAAACAATAAACTCAGT	2044
Qy	156	LysSerTyrIleLysGluSerValGlu-----AsnAspIle	169
		:::::	
Db	2045	TCAGAAAAAANAGTTGAGTAGGATGTTGGAGGTTTTTTTCTCTCAAAAAGAACTGTT	2104

```
Qy 170 LysPheAlaGlnGluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLysAsp 189
Db 2105 ---ATCCTTAAGAACATATTACTCAATTAGAAAAGAACTTCAGTTAATGGTTGAAGAC 2161
Qy 190 LeuGluSerIleLysValIleLysGluGlu----- 200
Db 2162 CAGATAATTAAATAAATGCTTGAATAATGAGCAAGTTCAGAACTTATTTGTTAAACT 2221
Qy 201 -----LysGluLysPheProSerSerProProThrThrProProSer 214
Db 2222 CAGTTGTATGGTTTCTTAAAGAAATG-----GGATCA 2254
Qy 215 ProAlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIle 234
Db 2255 GAAGTTTCAGAGACAGTGAAGAGAAAGAT----- 2284
Qy 235 GluThrLeuTyrAsnAsnLeuValAsnLysIleAspTyrLeuIleAsnLeuLysAla 254
Db 2285 -----GTTGTTAATGCTTCTCAGCAGTGGTGAATCTTG-----GCA 2323
Qy 255 LysIleAsnAsp-----CysAsnVal-----GluLysAspGluAlaHisValIle 270
Db 2324 AAAATAAATGAGAAATAATGCAACCTGCTTTTCAGCTGATGAAAAAGTATTAGAGTTA 2383
Qy 271 ThrLysLeuSerAspLeuLysAlaIleAspAspLysIleAspLeuPheLysAsnProTyr 290
Db 2384 GAAAA-----GAGATTAGTGCCTCAAGAA-----GAGAGTGTAGTT 2422
Qy 291 AspPheGluAlaIleLysLysLeuIleAsnAsp---AspThrLysLysAspMetLeugly 309
Db 2423 CAGTGTGAAGAACTTAACTTTATTGAGACACTATGAGCAAGAAAGTCTCTTAGG 2482
Qy 310 LysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThrIleIleSerLysLeuIle 329
Db 2483 AAAGAGTTAGAGAA-----ATACAGTCAGAAAAAGAGCCCTGCAGCTGTGATCTTCTA 2536
Qy 330 GluGlyLys-----PheGlnAspMetLeu----- 337
Db 2537 GAAATGAAGATGCTAATGAAAAACAAGGCTTGAATAATCAGAACTTTTAAATCAAGTT 2596
Qy 338 ---AsnIleSerGlnHisGlnCysValLysLysGlnCys---ProGluAsnSerGlyCys 355
Db 2597 GAAGAAGTATCTCAA---ACATGTAGCAAAAGTGAATCCATATGAAAAAGAAAAATGT 2653
Qy 356 PheArgHisLeuAspGluArgGluCysLysCysLysLeuLeuAsnTyrLysGlnGluGly 375
Db 2654 TTT-----ATAAAGGAACATGAAACCTAAAGCCACTACTAGAACAAAAAGAAATTACGA 2707
Qy 376 AspLysCysValGlu 380
Db 2708 GATAGGAGCAGAG 2722

RESULT 36
US-09-974-300-2643
; Sequence 2643, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE OF INVENTION: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2643
; LENGTH: 1678
; TYPE: DNA
```

```
; ORGANISM: Bacillus licheniformis
US-09-974-300-2643
```

Alignment Scores:

Pred. No.:	0.0223	Length:	1678
Score:	126.00	Matches:	91
Percent Similarity:	36.32%	Conservative:	63
Best Local Similarity:	21.46%	Mismatches:	128
Query Match:	5.54%	Indels:	142
DB:	10	Gaps:	21

US-10-057-531a-2 (1-431) x US-09-974-300-2643 (1-1678)

```
Qy 83 AlaGlyValTyrArgSerLeuLysLysGln-----IleGluLysAsnIle 97
Db 277 GCCGACAAATACCGTTTCAAAAATCGAAACAGGTCGCGCATATTGAGAACTTTTA 336
Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 337 TCAGCGGCTGAGTCCACATTTGAACATCTTA-----AAGGAA 375
Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
Db 376 ATCCCGCATCTTGTCTCAAGCAA-----GACGACAGCCGCAA 414
Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
Db 415 GAAATCGAAGAGGTAAGAAACGTTATACAAAAGTGCAGAAACCTTTTGGCATACAGC 474
Qy 158 TyrLysTyr-----IleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGly 175
Db 475 CACCTGTACGAGACCTCTATGCGAAAATTTGAAGCCGACCTCGATACCGTGTGGGAGGC 534
Qy 176 IleSerTyrTyrGlu-----LysValLeu 183
Db 535 ATTAACAGTTTGAAGAGAAACAGACGCGGAATTTACATAGAGCCAGAAAGTGTGTG 594
Qy 184 AlaLysTyrLysAspAspLeuGluSerIleLysLysValIleLysGluGluLysGluLys 203
Db 595 CTTGCCAAGACCGTTTGTGTTGAAGAGCTTCAATCATACATAGACGATGTGCGGAACATG 654
Qy 204 PheProSerSerProProThrThrProProSerProAlaLysThrAspGluGlnLysLys 223
Db 655 CTTGCCAGCTCCAGCAGACAGATTCCTCCGAGGAATTCGCAAAA----- 696
Qy 224 GluSerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuValAsn 243
Db 697 -----TTGAAAGCCGGCTATCAGGAATGATCGAT 726
Qy 244 Lys-----IleAspTyrIleuIleAsn 251
Db 727 AAAGGGTATAAGCTCGATCATATTCAAGTGGAGAGAGAGCTCGAAAACCTTTTGAAGAA 786
Qy 252 LeuLys-----AlaLysIleAsnAspCysAsnValGluLysAspGluAlaHis 267
Db 787 TTGAAGCCGCTGAAGATGCGCTTTTGAGCAGCTTGATTAGAGAAAGCCGCCCAATC 846
Qy 268 ValLysIleThrLysLeuSerAspLeuLysAlaIleAspAspLysIleAspLeuPheLys 287
Db 847 GTGCAGATC-----ATTGATGAACGATCCAGACGCTGTAT 882
Qy 288 AsnProTyrAspPheGlu----- 293
Db 883 AACCAGCTCGAGCATGAAGTCGAAGCAGGCGAGGAAATTTTGGGCAAAAGTCCTGAGCTT 942
Qy 294 -----AlaIleLysLysLeuIleAsnAspThrLysLysAsp----- 306
Db 943 GCCGCTGCATTTGAAAAAGCTT-----GAAGCCTCAAAAAGGATACCGAACGCTGAACACC 996
Qy 307 -----MetLeuGlyLysLeuLeuSerThrGlyLeuVal-----GlnAsnPhe 320
Db 997 GAGTCGTCAAGAAAGGCTACAGACTGACTACCGCGGAACCTGGAGAACGACGACGTATAT 1056
```

Qy 321 ProAsnThrIleSerLeuLeuLeuGluGlyLysPheGlnAsp----- 335
Db 1057 GAAAGCGCCTT-----GAAATGATCGAAAAGCAGTTCGAACAAGTGAAGAAAGGCTT 1110
Qy 336 -MetLeuAsnIleSerGlnHisGlnCysValLysLysGlnCysPro----- 350
Db 1111 GATCAGAAACATGTCGCTACTCGCTGTTAAAGAAAGAGCTTCGCGATATCGAAAGCAG 1170
Qy 351 -----GluAsnSerGly---CysPheArgHisLeuAspG1 361
Db 1171 ATGGAAGCGCGCAAGCGGAGCATGACGAAATACAGGACATGC----- 1213
Qy 361 uArgGluCysLysCys-----LeuLeuAsnTyrLysGlnGlu-GlyAspLysCysV 379
Db 1214 -----TCCAATGCTGAGAAAAGAGAGCTTCAGGCAAGAGAGCTTCTCAAGCAA 1263
Qy 379 alGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaThrC 399
Db 1264 TTGAA-----CGACACAAATCAAGGATACGGCGCGCAGCTCGAGAAAAGCAAT 1311
Qy 399 ysThrGluGluAspSerGlySerArgLysLysIleThrCys-----GluCysThrL 417
Db 1312 GTTCAGGCATCCCGAGGCCCAATACCGAAAAGATCCGACAGTCCGACAGCAGCTGCGAG 1371
Qy 417 ysProAsp 419
Db 1372 AAAGTGAC 1379

RESULT 37

US-09-880-107-2195
; Sequence 2195, Application US/09880107
; Patent NO. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2195
; LENGTH: 7573
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L33075
US-09-880-107-2195

Alignment Scores:
Pred. No.: 0.202 Length: 7573
Score: 125.50 Matches: 69
Percent Similarity: 37.97% Conservative: 81
Best Local Similarity: 17.47% Mismatches: 106
Query Match: 5.52% Indels: 139
Db: 10 Gaps: 16

US-10-057-531a-2 (1-431) x US-09-880-107-2195 (1-7573)

Qy 75 AspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGlu 94
Db 2889 GAAGTGTAAAGATTTCAGTCCCTGGCAAGAGTGCACCAAGCT---CGAAAGCGCTATCGA 2945
Qy 95 LysAsnIlePheThrPheAsnLeuAsnLeuAsnGlyLysPheGlnAsp----- 112
Db 2946 GATCCCTGCAGTACTTCGGGGACCATATAATGACATTATCAAAATCCAGGCTTTATT 3005

Qy 113 -----LysLysArgLysTyrPheLeuAspValLeuGluSerAsp----- 125
Db 3006 CGGCAAAACAAAGCTCGGAGTACTACAAGACTCTCTCATCAATGCTGAGAGCTCTCTCTATG 3065
Qy 126 -----LeuMetGlnPheLysHisIle-----SerSerAsnGluTyrIleIleGluAsp 141
Db 3086 GTTGTGGTCCGAAAATTTGTCCACCTGCTGGACCAAAAGTGCACCGAGTATTTTCAGGAGGAG 3125
Qy 142 SerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIle 161
Db 3126 CTTGACCTTATGAGATCGGGAAGAGGTTATCACCCCTCATCTCTCTCTCTCTCTCTCTCT 3173
Qy 162 LysGluSerValGluAsnAspIleLysPhe----- 171
Db 3174 AACAGCAGCTGGAGATGACCTCAATCTCATGGATATCAAAATTTGGACTGCTAGTGAAA 3233
Qy 172 -----AlaGlnGluGlyIleSerTyrTyrGluLysValLeuAlaLysTyrLys 187
Db 3234 AATAAGATTACGTTGCAAGGATGTGGTTTCCACAGTAAACAACTTACCAAAAAAATAAG 3293
Qy 188 AspAspLeuGlu-----SerIleLysLysValIle 197
Db 3294 GAACAGTTGCTGATATGATGATGATAAATAACAGAGGAGGTCTCAAGCTTTTGAGC 3353
Qy 198 LysGluLysLysGluLysPheProSerSerPro-----ProThrThr 211
Db 3354 AAGGAGAAGAGAGAGAGTTCGAAGCTTACCAGCAGCTGTTTTATTATTATGCAACCAAT 3413
Qy 212 ProProSerProAlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeu 231
Db 3414 CCCACCTATCTGGCCCAAGCTCATTTTTCAGATGCCCCAGAACAAAGTCCACCAAGTTTCATG 3473
Qy 232 ThrAsn---IleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIle 250
Db 3474 GACTCTGTAACTTTCACACTCTACAACTACGGCTCCACCCAGGAGAGGAGTACCTGCTC 3533
Qy 251 -----AsnLeuLysAlaLysIleAsnAspCysAsn 260
Db 3534 CTGCGGCTCTTTAAGACAGCAGCTCCCAAGAGGAAATCAAGTCAAGGTAGATCAGATTCAA 3593
Qy 261 -----ValGluLysAspGluAlaHisValLys----- 269
Db 3594 GAGATTGTGACAGAGAAATCCTACGGTTATTAAATGTTGTAAGTTTCAACCGTGGTGCC 3653
Qy 270 -----IleThrLysLysSerAspLeu 276
Db 3654 CGTGGCCAGATGCCCTGAGACAGATCTTGGCCCCAGCTCGTGAAGAAATTTATGATGATGAC 3713
Qy 277 LysAlaIleAsp-----AspLysIleAspLeuPheLysAsn----- 288
Db 3714 AAATCTCTCAACATCAAAATGACCCCTGGGATATTTACAAATCTTGGGTTAATCAGATG 3773
Qy 289 -----ProTyrAspPheGluAlaIleLysLysLeu 298
Db 3774 GAGTCTCAGACAGGAGAGGCAAGCAAACTGCCCTATGATGTGACCCCTGAGCAGCGCTA 3833
Qy 299 IleAsnAspAspThrLysLysAspMet----- 307
Db 3834 GCTCATGAAGAGTCAAGACAGCAGCGGTAGACAGCTCCATCAGGAACATCGGGCTGTGACA 3893
Qy 307 ----- 307
Db 3894 GACAAAGTTTCTCTCAGCCATTTGTCAGCTCTGTGGACAAAATCCCTTATGGGATGCGCTTC 3953
Qy 308 LeuGlyLysLeuLeuSerThrGlyLeuValGlnAsnPheProAsnThr----- 323
Db 3954 ATTGCCAAAGTGTCAAGGAGTCTGTTGATGAGAAAGTTTCCCTGATGCTGCTGGTGGATGAG 4013
Qy 324 -----IleIleSerLysLysLeuGlyLysPheGlnAsp 335
Db 4014 CTGCTGAAGATTATTGGTAACTTGTCTTTTATTATCATCATCAATGAAT 4058

RESULT 38

US-10-098-514-16
 ; Sequence 16, Application US/10098514
 ; Publication No. US20020194648A1
 ; GENERAL INFORMATION:

; APPLICANT: Chang, Sandra P
 ; APPLICANT: Christopher, David A
 ; APPLICANT: Vine, Benjamin
 ; APPLICANT: Su, Wei-wen
 ; APPLICANT: Bugos, Robert
 ; TITLE OF INVENTION: PLASMODIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PRODUCE
 ; FILE REFERENCE: A-71339/RET/TAL/NBC
 ; CURRENT APPLICATION NUMBER: US/10/098,514
 ; CURRENT FILING DATE: 2002-08-06
 ; PRIOR APPLICATION NUMBER: US 09/500,376
 ; PRIOR FILING DATE: 2000-02-08
 ; PRIOR APPLICATION NUMBER: US 60/274,599
 ; PRIOR FILING DATE: 2001-03-09
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 16
 ; LENGTH: 142

; TYPE: DNA
 ; ORGANISM: Artificial sequence
 ; FEATURE:
 ; OTHER INFORMATION: synthetic
 ; US-10-098-514-16

Alignment Scores:
 Pred. No.: 0.00893 Length: 142
 Score: 125.00 Matches: 21
 Percent Similarity: 76.09% Conservative: 14
 Best Local Similarity: 45.65% Mismatches: 11
 Query Match: 5.49% Indels: 0
 DB: 9 Gaps: 0

US-10-057-531A-2 (1-431) x US-10-098-514-16 (1-142)

QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
 Db 3 TTTCAAAATGTTTAAAGATCAGATTAAATTCATATAAAGATTAAACATCAAGTAATTAT 62
 QY 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
 Db 63 GTTGTCAAAGATCCATATAATTTCTTAATAAGAAAGAGATAAATTCCTTAAGCAGT 122

QY 158 TyrLysTyrIleLysGlu 163
 Db 123 TATAATTATTAAGGAT 140

RESULT 39

US-09-298-523B-70
 ; Sequence 70, Application US/09298523B
 ; Publication No. US20030059438A1
 ; GENERAL INFORMATION:

; APPLICANT: BRILES et al.
 ; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
 ; FILE REFERENCE: 454312-3140
 ; CURRENT APPLICATION NUMBER: US/09/298,523B
 ; CURRENT FILING DATE: 1999-04-23
 ; NUMBER OF SEQ ID NOS: 78
 ; SOFTWARE: PatentIn ver. 2.1
 ; SEQ ID NO 70
 ; LENGTH: 2405
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae

US-09-298-523B-70

Alignment Scores:
 Pred. No.: 0.0698 Length: 2405
 Score: 123.00 Matches: 81
 Percent Similarity: 38.30% Conservative: 45

Best Local Similarity: 24.62% Mismatches: 147
 Query Match: 5.41% Indels: 56
 DB: 9 Gaps: 12

US-10-057-531A-2 (1-431) x US-09-298-523B-70 (1-2405)

QY 19 LysGluThrAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeuGlyThr 38
 Db 973 AAAGAGCAAGGTAACCAAGAGCGGGCAACAGGAGGAGTTCTCGAGAGCTAGCAACA 1032
 QY 39 AspAsp-----AspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThr 56
 Db 1033 CTTGATAAAAGAAATGATCGCAAGTCTTCAGATTCTAGCGTAGGTGAAGAACTCTT 1092
 QY 57 MetAlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspVal 76
 Db 1093 CCAGCCCATCCCTTAATATGGCAAT---GAAAGTCACACAGACATAGGAAGATGTC 1149
 QY 77 ---IleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIleGluLys 95
 Db 1150 GATGAATATATAAAATAATGTTGAGT-----GAGATCCAATTAGATAGA 1194
 QY 96 AsnIlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArg 115
 Db 1195 AGAAACATACCCAAATGTCACTTAAACATAAAGTTGAGCGCAATTAAACGAAAGTAT 1254
 QY 116 LysTyrPheLeuAspValLeuGluSerAsp----- 125
 Db 1255 TTGTATGAATTAAAGTCTTTTAAAGAGAACTCGAAAAAGAGAGTTGAGTCATAAAACC 1314
 QY 126 -----LeuMetGlnPheLys 130
 Db 1315 AAAGCAGAGTTAAACCGCAGCTTTTGGACAGTTTAAATAAAGATACATTGAACCCAGAAAA 1374
 QY 131 HisIleSerSerAsnGluTyrIleIleGluAspSerPheLysLeuLeu----- 146
 Db 1375 AAGTAGCAGCAAGCTCGAAGAGAGTTGAAGAGTTGAAGAAATTAACCAAGCAAAAGAGTTGAG 1434
 QY 147 AsnSerGluGlnLysAsnThrLeuLeuLysSerTyrLysTyrIleLys---GluSerVal 165
 Db 1435 GAAGAAGATCGCGTAACCTACCAACCAATCTTACAAAAAGCTTCAACTTGAATTTGCT 1494
 QY 166 GluAsnAspIleLysPheAlaGlnGluGlyIleSerTyrTyrGluLysValLeuAlaLys 185
 Db 1495 GAGTCCGATGTGAAA-----GTTAAAAAGCGGAGCTTGAACCTAGTAAAA 1539
 QY 186 TyrLysAspLeuGluSerIleLysLysValIleLysGluGluLysGluLysPhePro 205
 Db 1540 GAGGAAGCTAACGAATCTCGAAGCAGGAGGAAAAATTAACCAAGCAAAAGAGAAAGTTGAG 1599
 QY 206 Ser---SerProThrThrProProSerProAlaLysThrAspGluGlnLysLysGlu 224
 Db 1600 ACTAAAAAGCTGAGGCTACAGGTTAGAAAAATCAACACAGATCGTAAAAAGCAGAA 1659
 QY 225 SerLysPheLeuProPheLeuThrAsnIleGluThrLeuTyrAsnAsnLeuValAsnLys 244
 Db 1660 GAAGAAGCTAAACGAAAAAGCAGAGAATCTGAGAAAAAGCTGCTGAAGCCCAACAAAAA 1719
 QY 245 IleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsnAspCysAsnValGluLysAsp 264
 Db 1720 GTGGATGCTGAAGATATATCTCTTGAAGCTAAATCGCTGAGTTGGAATATGAAAGTTGAG 1779
 QY 265 GluAlaHisValLysIleThrLysLeuSerAspLeuLysAlaIleAsp-----AspLys 282
 Db 1780 AGACTAGAA-----AAAGACTCAAGAGATTCAGTCTGACTCA 1821
 QY 283 IleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLysLeuIleAsnAspAsp 302
 Db 1822 GAAGATTTCTTAAAGAGAGCGCTCCGTGCTCTCTTCAATCTAAATTTG-----GAT 1872
 QY 303 ThrLysLysAspMetLeuGlyLysLeu 311
 Db 1873 ACCAAAAAGCTAAACTATCAAAACTT 1899

RESULT 40

US-09-298-523B-59
; Sequence 59, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 59
; LENGTH: 3463
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (319)..(3105)
; OTHER INFORMATION: coding sequence for PspA
US-09-298-523B-59

Alignment Scores:

Pred. No.: 0.116 Length: 3463
Score: 123.00 Matches: 81
Percent Similarity: 37.06% Conservative: 45
Best Local Similarity: 23.82% Mismatches: 148
Query Match: 5.41% Indels: 66
DB: 9 Gaps: 13

US-10-057-531A-2 (1-431) x US-09-298-523B-59 (1-3463)

QY 18 MetLysGluThrAlaAlaLysPheGluArgGlnHisMetAspSerPro----- 34
Db 1051 TTGAAGGAAGCTGTTGAAAAAGATGTAGCGACTTCAGACGAAGATAAACCAAGAGCGCG 1110
QY 35 -----AspLeuGlyThrAsp-----AspAspLysAlaMet 45
Db 1111 GCANAAACGAGGAGTTCTCGAGAGCTAGCAACACCTGATAAAGAAAGAAATATGCGCAAG 1170
QY 46 AlaAspIleGlySerIleGluGlyArgGlyThrMetAlaIleSerValThrMetAspAsn 65
Db 1171 TCTTCAGATTCTAGCGTAGGTGAAGAAACTCTTCCAGGCCATCCCTTAATATGCGCAAT 1230
QY 66 IleLeuSerGlyPheGluAsnGluTyrAspVal---IleTyrLeuLysProLeuAlaGly 84
Db 1231 ---GAAAGTCAGACAGACATAGGAAGATGTCGATGAATATATAAAAAAATGTTGAGT 1287
QY 85 ValTyrArgSerLeuLysLysGlnIleGluLysAsnIlePheThrPheAsnLeuLeu 104
Db 1288 -----GAGATCCCAATTAGTAGAAGAAACATACCCAAATGTCAACTTA 1332
QY 105 AsnAspIleLeuAsnSerArgLeuLysLysArgLysTyrPheLeuAspValLeuGluSer 124
Db 1333 AACATAAAGCTTGCGCAATTAAACGCAAGATTTCGTATGAATTAAGTGTGTTTAAAGAG 1392
QY 125 Asp----- 125
Db 1393 AACTCGAAAAAAGAGAGTTGAGTCGTCANAAACCAAGCAGAGTTAACCGCAGCTTTTGAG 1452
QY 126 -----LeuMetGlnPheLysHisIleSerSerAsnGluTyrIleIle 139
Db 1453 CAGTTTAAAAAGATACATTGAACCCAGAAAAAGAGTAGCAGACCTGAGAGAGAGGTT 1512
QY 140 GluAspSerPheLysLeuLeu-----AsnSerGluGlnLysAsnThrLeuLeu 155
Db 1513 GAAGAAGCTAAGAAAAAGCCAGGATCAAAAGAAAGAGATCGCGTAACCTACCCCAACC 1572
QY 156 LysSerTyrIleLys---GluSerValGluAsnAspIleLysPheAlaGlnGlu 174
Db 1573 AATACTTACAAAAACGCTTGAACCTTGAATTTGCTGAGTCGCGATGTGAAA----- 1620

QY 175 GlyIleSerTyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLys 194
Db 1621 ---GTTAAAGACGGAGCTTGAACCTAGTAAAGAGAGAGCTAACCAATCTCGAAACGAG 1677
QY 195 LysValIleLysGluGluLysGluLysPheProSer---SerProProThrThrProPro 213
Db 1678 GAAAAAATTAAAGCAAGCAAAAGAGAAAGTTGAGAGTAAAAAGCTCAGGCTACAAGTTA 1737
QY 214 SerProAlaLysThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsn 233
Db 1738 GAAAAAATCAAGACAGATCGTAAAAAAGCAGAGAAAGCTAAACGAAAAAGCAGAGAA 1797
QY 234 IleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLys 253
Db 1798 TCTGAGAAAAAGCTCTGCTGAAGCCAAACAAAGTGGATGCTGCTGAAGAATATGCTCTTGA 1857
QY 254 AlaLysIleAsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeu 273
Db 1858 GCTAAATCGCTGAGTTGGAATATGAAGTTTCAGAGACTAGAA----- 1899
QY 274 SerAspLeuLysAlaIleAsp-----AspLysIleAspLeuPheLysAsnProTyrAsp 291
Db 1900 AAAGAGCTCAAGAGAGATTGATGAGTCTGACTCAGAAAGATTATCTTAAAGAAAGCCTCCGT 1959
QY 292 PheGluAlaIleLysLysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeu 311
Db 1960 GCTCCTCTCAATCTAAATTG-----GATACCAAAAAAGCTAAACTATATCAAACTT 2010

Search completed: May 19, 2003, 17:01:06

Job time : 189 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 19, 2003, 15:02:45 ; Search time 1277 Seconds
(without alignments)
5466.135 Million cell updates/sec

Title: US-10-057-531A-2
Perfect score: 2275
Sequence: 1 MHHHHHSSGLVPRGSMKE.....TCECTKPDSPYPLDFGIFCSS 431

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16134066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_p2n.model -DEV=xlh
-O=/cnp2_1/USPTO.spool/US10057531/runat_12052003_091031_21676/app_query.fasta_1.583
-DB=EST -QPM=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10057531.ecgn_1.1.1456 @runat_12052003_091031_21676 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_oth:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	848	37.3	537	13	BI815756
2	736	32.4	419	14	BQ451367
3	618	27.2	402	13	BI670649
4	555	24.4	365	14	N97742
5	484	21.3	432	14	BQ633144
6	408	17.9	501	13	BI814631
7	403	17.7	422	14	BQ451709
8	398	17.5	522	14	BQ452445
9	393	17.3	455	14	BQ451036
10	317	13.9	303	13	BI815239
11	310.5	13.6	778	13	BI8164106
12	281	12.4	408	13	BI814931
13	281	12.4	418	14	BQ451572
14	281	12.4	422	14	BQ451661
15	279	12.3	595	13	BI677799
16	272	12.0	527	13	BI160302
17	272	12.0	784	13	BI166681
18	249.5	11.0	525	13	BI653325
19	243.5	10.7	439	13	BI160668
20	238.5	10.5	648	13	BI094652
21	238.5	10.5	661	13	BI094593
22	238.5	10.5	686	13	BI094577
23	234.5	10.3	415	13	BI094637
24	234.5	10.3	660	12	BF298872
25	233.5	10.3	405	13	BI094651
26	233	10.2	725	13	BI160032
27	229.5	10.1	542	13	BI167382
28	228.5	10.0	629	13	BI094603
29	225.5	9.9	571	13	BI094610
30	223	9.8	619	13	BI163150
31	203	8.9	782	13	BI164097
32	195	8.6	329	13	BI167649
33	192.5	8.5	624	13	BI094571
34	190.5	8.4	780	13	BI168979
35	189.5	8.3	693	13	BI094596
36	188	8.3	608	13	BI094595
37	184	8.1	652	13	BI166573
38	176.5	7.8	637	13	BI094585
39	166	7.3	362	14	BQ451205
40	164	7.2	686	13	BI167595
41	161	7.1	352	14	BQ577302
42	144.5	6.4	1067	17	CNS05PAD
43	144	6.3	558	13	BI094573
44	144	6.3	613	13	BI094586
45	133.5	5.9	699	9	AU061475

ALIGNMENTS

RESULT 1
BI815756
LOCUS
DEFINITION PfEST00a32c09.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:Q9TVG8 Q9TVG8 MSA1 PROTEIN ;, mRNA sequence.
ACCESSION BI815756
VERSION BI815756.1 GI:15909334
KEYWORDS EST.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 537)
AUTHORS Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,

Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I., Tsagaris, V., Belaygorod, L., Franklin, C., Carr, L., Grow, A., Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D., Waterston, R., Wilson, R., and Sibley, D.

TITLE JOURNAL COMMENT

WashU Plasmodium EST Project
Unpublished (2001)

Contact: L. David Sibley
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Library was constructed by Debopam Chakrabarti DNA sequencing by Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University

Seq primer: -40UP from Gibco

High quality sequence stop: 422.

FEATURES source

1. .537

Location/Qualifiers

/organism="Plasmodium falciparum"

/db_xref="taxon:5833"

/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"

/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"

/note="vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; Library was constructed by Debopam Chakrabarti.

Total RNA samples were isolated from mixed stage

saponin(0.1%) -lysed P. falciparum 3D7 infected

erythrocytes by the acidic guanidinium-phenol chloroform

method. The poly A+ RNA was isolated by the polyAT-tract

mRNA isolation system (Promega, WI) using streptavidin

Magnisphere particles. Directional cDNA libraries were

constructed by oligo d(T) priming of poly(A)+ RNA (5mg)

into EcoRI and XhoI sites of 1 ZapII vector using the Zap

cDNA synthesis kit (Stratagene, CA). The average size of the

cDNA inserts in the library was between 1.0 and 1.5kb.

Clones were mass excised using the ExAssist helper phage

(Stratagene), the phagemids were precipitated with PEG

8000 and extracted with phenol/chloroform. Phagemid DNA

was electroporated into DH10B cells."

217 a 75 c 83 g 162 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 4.84e-83 Length: 537

Score: 848.00 Matches: 154

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 37.27% Indels: 0

DB: 13 Gaps: 0

US-10-057-531a-2 (1-431) x B1815756 (1-537)

Qy 278 AlaileAspAspLysileAspLeuPheLysAsnProTyrAspPheGluAlaileLysLys 297

Db 2 GCATTTGATGACAAATAGATCTTTTAAACCCCTTACGACTTCGAGCAATTAATAAAA 61

Qy 298 LeuileAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317

Db 62 TTGATAATGATGATACGAAAAACATATGCTTGCAATTTACTTAGTACAGGATTAGTT 121

Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuileGluGlyLysPheClnAspMetLeu 337

Db 122 CAAAATTTTCTTAATACAAATATATCAAAATTAATTAAGGAAAAATTCACAGATATGTTA 181

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357

Db 182 ACATTTTCCACACACCAATGCGTAAAAAACCAATGTCAGAAAAATTCGATGTTTCAGA 241

Qy 358 HisLeuAspGluArgGluGlyCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377

Db 242 CATTAGATGAAGAGAGAGATGTAATGTTTATTAATTAACAAACAGAGGTGATATA 301

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaaspAla 397

Db 302 TGTGTTGAATCCAAATCCTACTTGTAAACGAATAATAGTGGATGTGATCGAGATGCC 361

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417

Db 362 ACATGTACCGAAGAAGATTTCAGGTAGCAGCAGAGAAATACATGTGAATGTACTAAA 421

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431

Db 422 CCGTATTCTTATCCACCTTTTCGATGTGATATTTCTGCAGTTC 463

RESULT 2

BO451367

LOCUS

DEFINITION

PFEST0aa98c03.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium

sequence.

BO451367

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,

Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,

Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I.,

Tsagaris, V., Belaygorod, L., Franklin, C., Carr, L., Grow, A.,

Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,

Waterston, R., Wilson, R., and Sibley, D.

WashU Plasmodium EST Project

Unpublished (2001)

CONTACT: L. David Sibley

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Library was constructed by Debopam Chakrabarti DNA sequencing by:

Washington University Genome Sequencing Center For information on

obtaining a clone please contact: L. David Sibley

(sibley@borcim.wustl.edu), Washington University

Seq primer: -40UP from Gibco.

FEATURES

source

1. .419

Location/Qualifiers

/organism="Plasmodium falciparum"

/db_xref="taxon:5833"

/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"

/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"

/note="vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; Library was constructed by Debopam Chakrabarti.

Total RNA samples were isolated from mixed stage

saponin(0.1%) -lysed P. falciparum 3D7 infected

erythrocytes by the acidic guanidinium-phenol chloroform

method. The poly A+ RNA was isolated by the polyAT-tract

mRNA isolation system (Promega, WI) using streptavidin

Magnisphere particles. Directional cDNA libraries were

constructed by oligo d(T) priming of poly(A)+ RNA (5mg)

into EcoRI and XhoI sites of 1 ZapII vector using the Zap

cDNA synthesis kit (Stratagene, CA). The average size of the

cDNA inserts in the library was between 1.0 and 1.5kb.

Clones were mass excised using the ExAssist helper phage

(Stratagene), the phagemids were precipitated with PEG

8000 and extracted with phenol/chloroform. Phagemid DNA

was electroporated into DH10B cells."

189 a 49 c 59 g 122 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 7.4e-71 Length: 419

Score: 736.00 Matches: 139
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 32.35% Indels: 0
 DB: 14 Gaps: 0

US-10-057-531A-2 (1-431) x B0451367 (1-419)

QY 237 LeuTyrAsnAsnLeuValAsnLysIleAspTyrLeuIleAsnLeuLysAlaLysIle 256
 DB 2 TTATACAAATAGTACTTAATAAATGACGATTACTTAATAAATGACGAAAGATT 61
 QY 257 AsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeu 276
 DB 62 AACGATTGTAATGTTGAAAGATGACACATGTTAAATAAATACTAACTAGTGATT 121
 QY 277 LysAlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaLys 296
 DB 122 AAAGCAATTTGATGACAAATAGATCTTTTAAACCCCTACGACTTCCGAAGCAAT 181
 QY 297 LysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeu 316
 DB 182 AAATTGATAAATGATGATACGAAAGATATGCTTGCGAAATTAATTAGTACAGATT 241
 QY 317 ValGlnAsnPheProAsnThrIleSerLysLeuIleGluLysPheGlnAspMet 336
 DB 242 GTTCAAAATTTTCTTAATAATAATATCAAAATTAATTAAGGAAATTTCCAAGAT 301
 QY 337 LeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPhe 356
 DB 302 TTAACATTTTCAACACCAATGCGTAAACAAACCAATGTCGAAATTTCTGATGTT 361
 QY 357 ArgHisLeuAspGluArgGluGluCysLysCysLeuAsnTyrLysGlnGluGly 375
 DB 362 AGACATTTAGATGAAGACAGAGATGTAATGTTTATTAATTAACAAAGGAT 418

RESULT 3

BI670649 402 bp mRNA linear EST 12-SEP-2001
 pfESToa02ell.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium
 falciparum cDNA 5' similar to TR:043995 043995 MEROZOITE SURFACE
 PROTEIN-1 ; mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 402)

Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,

Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,

Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I.,

Tsegareishvili, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A.,

Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,

Waterson, R., Wilson, R. and Sibley, D.

WashU Plasmodium EST Project

Unpublished (2001)

Contact: L. David Sibley

WashU Plasmodium EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

Library was constructed by Debopam Chakrabarti DNA sequencing by:

Washington University Genome Sequencing Center For information on

obtaining a clone please contact: L. David Sibley

(sibley@borcim.wustl.edu), Washington University

Seq primer: -40UP from Gibco

High quality sequence stop: 352.

Location/Qualifiers

1..402

/organism="Plasmodium falciparum"

FEATURES

source

/db_xref="taxon:5833"
 /clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
 /lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
 /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
 XhoI; Library was constructed by Debopam Chakrabarti.
 Total RNA samples were isolated from mixed stage
 saponin(0.1%)-lysed P. falciparum 3D7 infected
 erythrocytes by the acidic guanidinium-phenol chloroform
 method. The poly A+ RNA was isolated by the polyAT-Tract
 mRNA isolation system (Promega, WI) using streptavidin
 magnetic particles. Directional cDNA libraries were
 constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
 into EcoRI and XhoI sites of 1 ZapII vector using the Zap
 cDNA synthesis kit (Stratagene, CA). The average size of the
 cDNA inserts in the library was between 1.0 and 1.5kb.
 Clones were mass excised using the Exassist helper phage
 (Stratagene), the phagemids were precipitated with PEG
 8000 and extracted with phenol/chloroform. Phagemid DNA
 was electroporated into DH10B cells."

BASE COUNT 186 a 56 c 44 g 111 t 5 others

ORIGIN

Alignment Scores:

Pred. No.: 6.87e-58 Length: 402

Score: 618.00 Matches: 124

Percent Similarity: 94.03% Conservative: 2

Best Local Similarity: 92.54% Mismatches: 4

Query Match: 27.16% Indels: 4

DB: 13 Gaps: 1

US-10-057-531A-2 (1-431) x BI670649 (1-402)

QY 97 IlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLys 116

DB 1 ATTTTACATTTAAATTTAAATTTGAACGATATCTTAAATTCAGCTTTAAGAACCA 60

QY 117 TyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGlu 136

DB 61 TATTTCTTAGATCTATTAGCAATCTGATTTAATGCAATTTAAACATATATCCCAATGAA 120

QY 137 TyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLys 156

DB 121 TACATTTATTGAAGATTCATTTAAATTTATGCAATTCAGAACCAACACACACCTTTTAA 180

QY 157 SerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIle 176

DB 181 AGTTACAAATATATAAAGAAATCAGTAGAAATGATATTAAATTTCCAGGAAGGTATA 240

QY 177 SerTyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysVal 196

DB 241 AGTTATTATGAAAGGTTTTAGCGAAATATAAGGATGATTTAGCAATCAATTAAGGAT 300

QY 197 IleLysGluLysGluLysPheProSerSerProThrThrProProSerProAla 216

DB 301 ATCAAGAGAGAAAGGAGAGATTCATCATCACCACCAACACACCTCCGCACGACGC 360

QY 217 LysThrAsp-----GluGlnLysLysGluSerLys 226

DB 361 NNNNAACCAACCCCAACATCCCAACCAACCAACCAACCAACCAACCAACCAACCA 402

RESULT 4

N97742

LOCUS

DEFINITION

1183C3 czappFDd2.1, Debopam Chakrabarti Plasmodium falciparum cDNA

clone PF1183C, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Plasmodium falciparum

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE

1 (bases 1 to 365)

AUTHORS

Dame, J.B., Arnot, D.E., Bourke, P., Chakrabarti, D., Christodoulou, Z.,

N97742 365 bp mRNA linear EST 18-NOV-1996

1183C3 czappFDd2.1, Debopam Chakrabarti Plasmodium falciparum cDNA

clone PF1183C, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Plasmodium falciparum

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE

1 (bases 1 to 365)

AUTHORS

Dame, J.B., Arnot, D.E., Bourke, P., Chakrabarti, D., Christodoulou, Z.,

N97742 365 bp mRNA linear EST 18-NOV-1996

1183C3 czappFDd2.1, Debopam Chakrabarti Plasmodium falciparum cDNA

clone PF1183C, mRNA sequence.

Coppel, R., Cowman, A., Craig, A., Fischer, K., Foster, J., Goodman, N., Hinterberg, K., Holder, A.A., Holt, D., Kemp, D., Lanzer, M., Lim, A., Newbold, C., Ravetch, J.V., Reddy, G.R., Rubio, J., Schuster, S.M., Su, X.-Z., Thompson, J.K., Vital, F., Wellens, T.E. and Werner, E.
Current status of the Plasmodium falciparum genome project
Mol. Biochem. Parasitol. 79, 1-12 (1996)
97001675

COMMENT
Contact: Debopam Chakrabarti
Department of Molecular Biology and Microbiology
University of Central Florida
Orlando, FL 32816-2360
Tel: 407 384 2061
Fax: 407 384 3095
Email: dchak@pegasus.cc.ucf.edu
Seq primer: T3.

FEATURES

source

1. 365
Location/Qualifiers
/organism="Plasmodium falciparum"
/strain="Dd2"
/db_xref="taxon:5833"
/clone="PF1183C"
/lab_host="E. coli XL-1 blue"
/note="Vector: lambda ZAP II; Site_1: EcoR I; Site_2: Xho I; PolyA+ RNA, from asynchronous blood stage parasites of the Dd2 isolate cultured in vitro, was reverse transcribed using an oligo dT-Xho I primer. Second strand was prepared using RNase H and DNA polymerase I. EcoR I adapters were ligated to the cDNA, and it was digested with Xho I. Prepared fragments were ligated into EcoR I + Xho I digested lambda ZAP II vector."
163 a 39 c 43 g 120 t

BASE COUNT
ORIGIN

Alignment Scores:

Pred. No.: 5,2e-51 Length: 365
Score: 555.00 Matches: 111
Percent Similarity: 99.11% Conservations: 0
Best Local Similarity: 99.11% Mismatches: 1
Query Match: 24.40% Indels: 0
DB: 14 Gaps: 0

US-10-057-531A-2 (1-431) x N97742 (1-365)

Qy 58 AlarSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValle 77
|||||
Db 28 GCAATATCTGTCAACATGGATAATATCTCTCAGGATTTGAAATGAATATGATGTTATA 87
|||||
Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
|||||
Db 88 TATTTAAACCTTTAGCTGGAGTATATAGAACCTTAAACAAATTTGAAAAAACATT 147
|||||
Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
|||||
Db 148 ATTACATTTAATTATTTTGAACGATATCTTAATTCACGCTTTAGAACGAAATAT 207
|||||
Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
|||||
Db 208 TTCTTAGATGATTAGAATCTGATTTAATGAATTTAAACATATATCTCAATGAATAC 267
|||||
Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
|||||
Db 268 ATTATTGAAGATTCATTTAAATTTAATTCAGAACAAAAACACACTTTTAAAAAGT 327
|||||
Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIle 169
|||||
Db 328 TACAATATATAAAGAAATCATAGTAAAAATGATATT 363
|||||

RESULT 5
BQ633144 432 bp mRNA linear EST 03-JUL-2002
LOCUS
DEFINITION PFESToab38c11.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ;

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

FEATURES
source

mRNA sequence.
BQ633144
BQ633144.1 GI:21687297
EST.
malaria parasite P. falciparum.
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 432)
Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I., Tsagarishvili, R., Belaygorod, L., Franklin, C., Carr, L., Grov, A., Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D., Waterston, R., Wilson, R. and Sibley, D.
WashU Plasmodium EST Project
Unpublished (2001)
Contact: L. David Sibley
WashU Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Library was constructed by Debopam Chakrabarti DNA sequencing by:
Washington University Genome Sequencing Center for information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Seq primer: -400P from Gibco.
Location/Qualifiers
1. 432
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; Library was constructed by Debopam Chakrabarti.
Total RNA samples were isolated from mixed stage saponin(0.1%)-lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyA+-Tract mRNA isolation system (Promega, WI) using streptavidin Magnesphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 ZapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the ExAssist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

BASE COUNT 166 a 49 c 59 g 158 t

ORIGIN

Alignment Scores:
Pred. No.: 4,29e-43 Length: 432
Score: 484.00 Matches: 83
Percent Similarity: 100.00% Conservations: 1
Best Local Similarity: 98.81% Mismatches: 0
Query Match: 21.27% Indels: 0
DB: 14 Gaps: 0

US-10-057-531A-2 (1-431) x BQ633144 (1-432)

Qy 348 GlnCysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCys 367
|||||
Db 3 CAATGTCAGAAAATTCGGATGTTTCAGACATTTAGTGAAGAGAGAGATGTAATGT 62
|||||
Qy 368 LeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsnProThrCysAsn 387
|||||
Db 63 TTATTAATTTACAACCAAGAGGTGATAAATGTCTTGAATAATCCAAATCTCTACTGTGTAAC 122
|||||
Qy 388 GluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGluGluAspSerGlySerSer 407
|||||

```

Db 123 GAAAAAATGGTGATGTCAGATGCCACATGATCCAGAAAGATTCAGTACGAC 182
QY 408 ArGlyLysIleThrCysGluCysThrLysProAspSerTyrProLeuPheAspGlyIle 427
|||||
Db 183 AGAAAGAAATACATGCTGAATGTTCTAAACCTGATCTTATCCACTTTTCGATGTATT 242
QY 428 PheCysSerSer 431
|||||
Db 243 TTCTGCGATGCC 254

RESULT 6
BI814631/c
LOCUS
DEFINITION
PFEStoaa40c07.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium
falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ;
mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
Marra,M., Hillier,L., Martin,J., Wyllie,T., Dente,M., Theising,B.,
Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
Tsagarrelshvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,
Waterston,R., Wilson,R. and Sibley,D.
WashU Plasmodium EST Project
Unpublished (2001)
Contact: L. David Sibley
WashU Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library was constructed by Debopam Chakrabarti DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Possible reversed clone: similarity on wrong strand
Seq primer: -400P from Gibco
High quality sequence stop: 261.
Location/Qualifiers
1..501
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
/lab_host="DH10B (GensHog, Invitrogen, Inc.)"
/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
XhoI; Library was constructed by Debopam Chakrabarti.
Total RNA samples were isolated from mixed stage
saponin(0.1%) lysed P. falciparum 3D7 infected
erythrocytes by the acidic guanidinium-phenol chloroform
method. The poly A+ RNA was isolated by the polyAT-Tract
mRNA isolation system (Promega, WI) using streptavidin
Magnesphere particles. Directional cDNA libraries were
constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
into EcoRI and XhoI sites of 1 zapII vector using the Zap
cDNA synthesis kit (Stragene, CA). The average size of the
cDNA inserts in the library was between 1.0 and 1.5kb.
Clones were mass excised using the EXassist helper phage
(Stratagene), the phagemids were precipitated with PEG
8000 and extracted with phenol/chloroform. Phagemid DNA
was electroporated into DH10B cells."
181 a 62 c 51 g 207 t

BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.: 1.22e-34 Length: 501
Score: 408.00 Matches: 71

```

```

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.93% Indels: 0
DB: 13 Gaps: 0
US-10-057-531A-2 (1-431) x BI814631 (1-501)

QY 361 GluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGlu 380
|||||
Db 501 GAAAGAGAGAGATGTAATGTTTATTAAATTACAACACAGAGGTGATAATGTGTGAA 442
QY 381 AsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAlaThrCysThr 400
|||||
Db 441 AATCCAAATCCTACTTGTGTAACGAAATAATGGTGGATGTGATGCAGATGCCATGTACC 382
QY 401 GluGluAspSerCysSerArgLysLysIleThrCysGluCysThrLysProAspSer 420
|||||
Db 381 GAAGAAGATTTCAGGTAGCAGACAGAAAGAAATCAGATGTGAATGTACTAAACCTGATTCT 322
QY 421 TyrProLeuPheAspGlyIlePheCysSerSer 431
|||||
Db 321 TATCCACTTTTCGATGGTATTTTCTGCGATGCC 289

RESULT 7
BQ451709
LOCUS
DEFINITION
PFEStoab07h03.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium
falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ;
mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
Marra,M., Hillier,L., Martin,J., Wyllie,T., Dente,M., Theising,B.,
Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
Tsagarrelshvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,
Waterston,R., Wilson,R. and Sibley,D.
WashU Plasmodium EST Project
Unpublished (2001)
Contact: L. David Sibley
WashU Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library was constructed by Debopam Chakrabarti DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Seq primer: -400P from Gibco.
Location/Qualifiers
1..422
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
/lab_host="DH10B (GensHog, Invitrogen, Inc.)"
/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
XhoI; Library was constructed by Debopam Chakrabarti.
Total RNA samples were isolated from mixed stage
saponin(0.1%) lysed P. falciparum 3D7 infected
erythrocytes by the acidic guanidinium-phenol chloroform
method. The poly A+ RNA was isolated by the polyAT-Tract
mRNA isolation system (Promega, WI) using streptavidin
Magnesphere particles. Directional cDNA libraries were
constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
into EcoRI and XhoI sites of 1 zapII vector using the Zap
cDNA synthesis kit (Stragene, CA). The average size of the
cDNA inserts in the library was between 1.0 and 1.5kb.
Clones were mass excised using the EXassist helper phage
(Stratagene), the phagemids were precipitated with PEG
8000 and extracted with phenol/chloroform. Phagemid DNA
was electroporated into DH10B cells."
181 a 62 c 51 g 207 t

BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.: 1.22e-34 Length: 501
Score: 408.00 Matches: 71

```

cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the EXAssist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

BASE COUNT 162 a 44 c 57 g 159 t

ORIGIN

Alignment Scores:
 Pred. No.: 3.45e-34 Length: 422
 Score: 403.00 Matches: 70
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 17.71% Indels: 0
 DB: 14 Gaps: 0

US-10-057-531A-2 (1-431) x B0451709 (1-422)

Qy 362 ArgGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsn 381
 |||||
 Db 3 AGAGAAGAAATGTAATGTTTAAATACAAACAGAGGTGATAAATGTTGAAAT 62
 |||||
 Qy 382 ProAsnProThrCysAsnGluAsnGlyCysAspAlaAspAlaThrCysThrGlu 401
 |||||
 Db 63 CCAATCTACTTGAACGAAATTAATGGTGGATGTCAGATGCCACATGTACCGAA 122
 |||||
 Qy 402 GluAspSerGlySerArgLysLysIleThrCysGluCysThrLysProAspSerTyr 421
 |||||
 Db 123 GAAGATTCAGTACGACAGACAGAAATACATGTGATGTAACCTGATTCATAT 182
 |||||
 Qy 422 ProLeupheAspGlyIlePheCysSerSer 431
 |||||
 Db 183 CCATCTTCGATGGTATTTCTTCGAGTTC 212
 |||||

RESULT 8

B0452445 522 bp mRNA linear EST 29-MAY-2002
 LOCUS PFESToab03a10.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium
 DEFINITION falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ;
 mRNA sequence.

ACCESSION B0452445

VERSION B0452445.1 GI:21255557

KEYWORDS EST.

SOURCE malaria parasite P. falciparum.

ORGANISM Plasmodium falciparum

REFERENCE 1 (bases 1 to 522)

AUTHORS Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,
 Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,
 Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I.,
 Tsagareishvili, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A.,
 Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,
 Waterston, R., Wilson, R. and Sibley, D.

TITLE WashU Plasmodium EST Project

JOURNAL Unpublished (2001)

COMMENT

Contact: L. David Sibley
 WashU Plasmodium EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: est@wustl.wustl.edu
 Library was constructed by Debopam Chakrabarti DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: L. David Sibley
 (sibley@borcim.wustl.edu), Washington University
 Seq primer: -400P from Gibco
 High quality sequence stop: 422.

FEATURES

source

1. .522

/organism="Plasmodium falciparum"

/db_xref="taxon:5833"

/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"

/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
 /note="Vector: pBluescript SK plus; Site 1: EcoRI; Site 2:
 XhoI; Library was constructed by Debopam Chakrabarti.

Total RNA samples were isolated from mixed stage
 saponin(0.1%) lysed P. falciparum 3D7 infected
 erythrocytes by the acidic guanidinium-phenol chloroform
 method. The poly A+ RNA was isolated by the polyAT-Tract
 mRNA isolation system (Promega, WI) using streptavidin
 Magnosphere particles. Directional cDNA libraries were
 constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
 into EcoRI and XhoI sites of 1 ZapII vector using the Zap
 cDNA synthesis kit (Stratagene, CA). The average size of the
 cDNA inserts in the library was between 1.0 and 1.5kb.
 Clones were mass excised using the EXAssist helper phage
 (Stratagene), the phagemids were precipitated with PEG
 8000 and extracted with phenol/chloroform. Phagemid DNA
 was electroporated into DH10B cells."

BASE COUNT 204 a 63 c 68 g 187 t

ORIGIN

Alignment Scores:

Pred. No.: 1.63e-33 Length: 522
 Score: 398.00 Matches: 69
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 17.49% Indels: 0
 DB: 14 Gaps: 0

US-10-057-531A-2 (1-431) x B0452445 (1-522)

Qy 363 GluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnPro 382
 |||||
 Db 1 GAAGAATGTAATGTTTAAATACAAACAGAGGTGATAAATGTTGAAATCCA 60
 |||||

Qy 383 AsnProThrCysAsnGluAsnGlyCysAspAlaAspAlaThrCysThrGluGlu 402
 |||||

Db 61 AATCCTACTTGTAAACGAAATTAATGGTGGATGTCAGATGCCACATGTACCGAAGAA 120
 |||||

Qy 403 AspSerGlySerSerArgLysLysIleThrCysGluCysThrLysProAspSerTyrPro 422
 |||||

Db 121 GATTTCAGTACGACAGACAGAAATACATGTGATGTAACCTGATTCATATCCA 180
 |||||

Qy 423 LeupheAspGlyIlePheCysSerSer 431
 |||||

Db 181 CTTTTCGATGGTATTTCTTCGAGTTC 207
 |||||

RESULT 9

B0451036 455 bp mRNA linear EST 29-MAY-2002
 LOCUS PFESToab03a10.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium
 DEFINITION falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ;
 mRNA sequence.

ACCESSION B0451036

VERSION B0451036.1 GI:21254148

KEYWORDS EST.

SOURCE malaria parasite P. falciparum.

ORGANISM Plasmodium falciparum

REFERENCE 1 (bases 1 to 455)

AUTHORS Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,
 Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,
 Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I.,
 Tsagareishvili, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A.,
 Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,
 Waterston, R., Wilson, R. and Sibley, D.

TITLE WashU Plasmodium EST Project

JOURNAL Unpublished (2001)

COMMENT

Contact: L. David Sibley
 WashU Plasmodium EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: estwatson.wustl.edu
Library was constructed by Debopam Chakrabarti DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Seq primer: -400P from Gibco
High quality sequence stop: 426.

FEATURES

Location/Qualifiers
1..455
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
Total RNA samples were isolated from mixed stage
saponin(0.1%) lysed P. falciparum 3D7 infected
erythrocytes by the acidic guanidium-phenol chloroform
method. The poly A+ RNA was isolated by the polyAT-tract
mRNA isolation system (Promega, WI) using streptavidin
Magnesphere particles. Directional cDNA libraries were
constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
into EcoRI and XhoI sites of 1 ZapII vector using the Zap
cDNA synthesis kit (Stratagene, CA). The average size of the
cDNA inserts in the library was between 1.0 and 1.5kb.
Clones were mass excised using the ExAssist helper phage
(Stratagene), the phagemids were precipitated with PEG
8000 and extracted with phenol/chloroform. Phagemid DNA
was electroporated into DH10B cells."

BASE COUNT 179 a 46 c 56 g 174 t
ORIGIN
Alignment Scores:
Pred. No.: 4..82e-33 Length: 455
Score: 393.00 Matches: 68
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.2% Indels: 0
DB: 14 Gaps: 0
US-10-057-531A-2 (1-431) x BQ451036 (1-455)

QY 364 GluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLysCysValGluAsnProAsn 393
Db 3 GAATGTAATGTTTATTAAATTAACAACAGAGGTTGTTGTAATAATCCAAAT 62
QY 384 ProThrCysAsnGluAsnGlyGlyCysAspAlaAspAlaThrCysThrGluGluAsp 403
Db 63 CCTACTGTACCAAAATAATGTTGGATGTGATGCAGATGCCACATGTACCGAAGAGAT 122
QY 404 SerGlySerSerArgLysLysIleThrCysGluCysThrLysProAspSerTyrProLeu 423
Db 123 TCAGGTAGCAGCAGAAAGAAATACATGTGAATGTACTAAACCTGATTTTATCCACTT 182
QY 424 PheAspGlyIlePheCysSerSer 431
Db 183 TTCGATGGTATTTTCTGCAGTTCC 206

RESULT 10
LOCUS BI815239 303 bp mRNA linear EST 03-OCT-2001
DEFINITION PfEST0016b11.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ; mRNA sequence.
ACCESSION BI815239
VERSION BI815239.1 GI:15908372
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 303)
AUTHORS Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,

TITLE JOURNAL COMMENT

Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I., Tsagarelshvili,R., Belaygorod,L., Franklin,C., Carrillo, G.W.A., Maquire,L., Richey,J., Wadhwa,J., Kennedy,S., Levinso,D., Waterston,R., Wilson,R. and Sibley,D.
WashU Plasmodium EST Project
Unpublished (2001)
Contact: L. David Sibley
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estwatson.wustl.edu
Library was constructed by Debopam Chakrabarti DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Putative full length read
vector to vector length is 433
Seq primer: -400P from Gibco
High quality sequence stop: 257.

FEATURES

Location/Qualifiers
1..303
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
XhoI; Library was constructed by Debopam Chakrabarti.
Total RNA samples were isolated from mixed stage
saponin(0.1%) lysed P. falciparum 3D7 infected
erythrocytes by the acidic guanidium-phenol chloroform
method. The poly A+ RNA was isolated by the polyAT-tract
mRNA isolation system (Promega, WI) using streptavidin
Magnesphere particles. Directional cDNA libraries were
constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
into EcoRI and XhoI sites of 1 ZapII vector using the Zap
cDNA synthesis kit (Stratagene, CA). The average size of the
cDNA inserts in the library was between 1.0 and 1.5kb.
Clones were mass excised using the ExAssist helper phage
(Stratagene), the phagemids were precipitated with PEG
8000 and extracted with phenol/chloroform. Phagemid DNA
was electroporated into DH10B cells."

BASE COUNT 106 a 41 c 41 g 115 t
ORIGIN
Alignment Scores:
Pred. No.: 6..47e-25 Length: 303
Score: 317.00 Matches: 55
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 13.93% Indels: 0
DB: 13 Gaps: 0
US-10-057-531A-2 (1-431) x BI815239 (1-303)

QY 377 LysCysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAsp 396
Db 3 AAATGTTGAAATCCAAATCTACTTGTACGAAATAATGTTGGATGTGATGCAGAT 62
QY 397 AlaThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThr 416
Db 63 GCCACATGTACCGAAGAGATTTCAGGTAGCAGCAGAGAAAGAAATACATGTGAATGTACT 122
QY 417 LysProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
Db 123 AAACCTGATTCTTATCCACTTTTCGATGGTATTTCCTGCAGTTCC 167
RESULT 11
LOCUS BM164106 778 bp mRNA linear EST 04-DEC-2001
DEFINITION EST566629 PyBS Plasmodium yoelii cDNA clone PYCUP75 5' end, mRNA sequence.

Washington University Genome Sequencing Center For information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Putative full length read
vector to vector length is 409
Seq primer: -40UP from Gibco.

FEATURES

source
1. 408
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
XhoI; Library was constructed by Debopam Chakrabarti.
Total RNA samples were isolated from mixed stage
saponin(0.1%) lysed P. falciparum 3D7 infected
erythrocytes by the acidic guanidinium-phenol chloroform
method. The poly A+ RNA was isolated by the polyAT-Tract
mRNA isolation system (Promega, WI) using streptavidin
Magnesphere particles. Directional cDNA libraries were
constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
into EcoRI and XhoI sites of 1 ZapII vector using the Zap
cDNA synthesis kit (Stratagene, CA). The average size of the
cDNA inserts in the library was between 1.0 and 1.5kb.
Clones were mass excised using the ExAssist helper phage
(Stratagene), the phagemids were precipitated with PEG
8000 and extracted with phenol/chloroform. Phagemid DNA
was electroporated into DH10B cells."

BASE COUNT 161 a 43 c 46 g 158 t
ORIGIN
Alignment Scores:
Pred. No.: 8.9e-21 Length: 408
Score: 281.00 Matches: 49
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.35% Indels: 0
DB: 13 Gaps: 0

US-10-057-531A-2 (1-431) x B1814931 (1-408)

QY 383 AsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGluGlu 402
|||||
Db 3 AATCCTACTTGTAAACGAAATAATGTTGGATGTGATGCAGATGCCACATGTACCGAAGAA 62
QY 403 AspSerGlySerArgLysLysIleThrCysGluCysThrLysProAspSerTyrPro 422
|||||
Db 63 GATTCAGGTAGCAGCAGAGAAATAATACATGTGATGTACTAAACCTGATCTTATCCA 122
QY 423 LeuPheAspGlyIlePheCysSerSer 431
|||||
Db 123 CTTTTCGATGGTATTTTCTGCAGTTCC 149

RESULT 13
BQ451572 418 bp mRNA linear EST 29-MAY-2002
LOCUS pFESToab06b11.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium
falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ;
mRNA sequence.

ACCESSION BQ451572
VERSION BQ451572.1 GI:21254684
KEYWORDS EST.
SOURCE malaria parasite P. falciparum.

ORGANISM Plasmodium falciparum
DEFINITION Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 418)

AUTHORS Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
Marra,M., Hillier,L., Martin,J., Wyllie,T., Dente,M., Theisling,B.,
Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
Tsagarashvili,L., R., Belaygorod,L., Franklin,C., Carr,L., Graw,A.,
Maquire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,
Waterston,R., Wilson,R. and Sibley,D.
WashU Plasmodium EST Project

TITLE

JOURNAL COMMENT

Unpublished (2001)
Contact: L. David Sibley
WashU Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library was constructed by Debopam Chakrabarti DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Seq primer: -40UP from Gibco.

FEATURES

source

1. 418
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
XhoI; Library was constructed by Debopam Chakrabarti.
Total RNA samples were isolated from mixed stage
saponin(0.1%) lysed P. falciparum 3D7 infected
erythrocytes by the acidic guanidinium-phenol chloroform
method. The poly A+ RNA was isolated by the polyAT-Tract
mRNA isolation system (Promega, WI) using streptavidin
Magnesphere particles. Directional cDNA libraries were
constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
into EcoRI and XhoI sites of 1 ZapII vector using the Zap
cDNA synthesis kit (Stratagene, CA). The average size of the
cDNA inserts in the library was between 1.0 and 1.5kb.
Clones were mass excised using the ExAssist helper phage
(Stratagene), the phagemids were precipitated with PEG
8000 and extracted with phenol/chloroform. Phagemid DNA
was electroporated into DH10B cells."

BASE COUNT 171 a 43 c 46 g 158 t
ORIGIN

Alignment Scores:
Pred. No.: 9.2e-21 Length: 418
Score: 281.00 Matches: 49
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.35% Indels: 0
DB: 14 Gaps: 0

US-10-057-531A-2 (1-431) x BQ451572 (1-418)

QY 383 AsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGluGlu 402
|||||
Db 3 AATCCTACTTGTAAACGAAATAATGTTGGATGTGATGCAGATGCCACATGTACCGAAGAA 62
QY 403 AspSerGlySerArgLysLysIleThrCysGluCysThrLysProAspSerTyrPro 422
|||||
Db 63 GATTCAGGTAGCAGCAGAGAAATAATACATGTGATGTACTAAACCTGATCTTATCCA 122
QY 423 LeuPheAspGlyIlePheCysSerSer 431
|||||
Db 123 CTTTTCGATGGTATTTTCTGCAGTTCC 149

RESULT 14
BQ451661

LOCUS pFESToab07c09.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium
falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ;
mRNA sequence.

ACCESSION BQ451661
VERSION BQ451661.1 GI:21254773
KEYWORDS EST.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
DEFINITION Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 422)

AUTHORS Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I., Tsagareishvili, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A., Maquire, L., Richey, J., Watkins, J., Kennedy, S., Levinso, D., Waterston, R., Wilson, R. and Sibley, D.

TITLE WashU Plasmodium EST Project

JOURNAL Unpublished (2001)

COMMENT Contact: L. David Sibley
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library was constructed by Debopam Chakrabarti DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Seq primer: -40UP from Gibco.

FEATURES Location/Qualifiers

1..422

source

1..422

/organism="Plasmodium falciparum"

/db_xref="taxon:5833"

/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"

/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"

/note="vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage saponin(0.1%) lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-tract mRNA isolation system (Promega, WI) using streptavidin Magnosphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 ZapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the EXassist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

BASE COUNT 175 a 43 c 46 g 158 t

ORIGIN

Alignment Scores:

Pred. No.: 9.32e-21 Length: 422

Score: 281.00 Matches: 49

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 12.35% Indels: 0

DB: 14 Gaps: 0

US-10-057-531A-2 (1-431) x BQ451661 (1-422)

Qy 383 AsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAlaThrCysThrGluGlu 402

Db 3 AATCTACTTGTACGAAATAAATGCTGATGTGATGTCAGATGCCACATGTACCGAAGA 62

Qy 403 AspSerGlySerSerArgLysLysIleThrCysGluCysThrLysProAspSerTyrPro 422

Db 63 GATTTCAGTACGACGAGAAAGAAATACATGTCGATGTAATGCTACTAAACCTGCTTATCCA 122

Qy 423 LeuPheAspGlyIlePheCysSerSer 431

Db 123 CTTTTCGATGGTATTTCTGCAGTTCC 149

RESULT 15

LOCUS BM167799 595 bp mRNA linear EST 04-DEC-2001

DEFINITION EST570322 PYBS Plasmodium yoelii yoelii cDNA clone PYCOQ81 5' end, mRNA sequence.

ACCESSION BM167799

VERSION BM167799.1 GI:17301031

KEYWORDS EST.

SOURCE Plasmodium yoelii yoelii.

ORGANISM Plasmodium yoelii yoelii

REFERENCE Eukaryota, Alveolata; Apicomplexa; Haemosporida; Plasmodium.

AUTHORS 1 (bases 1 to 595)
Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B., Fraser, C.M. and Carucci, D.J.

TITLE Plasmodium yoelii EST project at TIGR

JOURNAL Unpublished (2001)

COMMENT Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

FEATURES Location/Qualifiers

1..595

source

1..595

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCOQ81"

/clone_lib="PYBS"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="vector: pAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybrizAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybrizAP vector and plasmid DNA isolated."

BASE COUNT 254 a 64 c 111 g 166 t

ORIGIN

Alignment Scores:

Pred. No.: 2.47e-20 Length: 595

Score: 279.00 Matches: 70

Percent Similarity: 54.85% Conservative: 43

Best Local Similarity: 33.98% Mismatches: 78

Query Match: 12.26% Indels: 16

DB: 13 Gaps: 7

US-10-057-531A-2 (1-431) x BM167799 (1-595)

Qy 231 LeuThrAsnIleGluThrLeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIle 250

Db 5 CTTGAAGATCTTAAAGGATTATTAAGAACCCGTAATAGTCAAGCAAGAATATTCAGAA 64

Qy 251 AsnLeuLysAlaLysIleAsnAsnCysAsnValGluLysAspGluAlaHisValLysIle 270

Db 65 GAATTACAAATAGACTTGATTAATAAATGAAAAAGCTGAATTTGAATATTATACA 124

Qy 271 ThrLysLeuSerAspLeuLysAlaIleAspLysIleAspLysIleAspLeuPheLysAsnProTyr 290

Db 125 AAAAATTTAGAAAAATACATAAAATTCACGAAAAACCTGACGAATTTGTAGAACATCCA 184

Qy 291 AspPheGluAlaIleLysLysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLys 310

Db 65 GAATTACAAATAGACTTCGATAATTATAAAAATGAAAAACCTGAATTTTGAAATATTAACA 124

QY 271 ThrLysLeuSerAspLeuLysAlaIleAspAspLysIleAspLeuPheLysAsnProTyr 290
 Db 125 AAAAAATTTAGAAAAATACATACAAATGACGAAATCTGACCAATTTGTAGAACATGCA 184
 QY 291 AspPheGluAlaIleLysLysLeuLysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLys 310
 Db 185 GAA-----AATAATAAACACATAGCCCTCAATAGCTTTAAACAAC 223
 QY 311 LeuLeuSerThrGlyLeuVal---GlnAsnProAsnThrIleIleSerLysLeuIle 329
 Db 224 TTAATAATAATCTGTTAGTAGAGAGAGGTGAATCAAGAAATAATATGACAAAATGCTT 283
 QY 330 GluGlyLysPheGlnAspMetLeuAsnIleSer---GlnHisGlnCysVal---LysLys 347
 Db 284 ACATCGATGCTATGATTTATAGGTAGACCCCTAACATGATGTGTGTATACAGA 343
 QY 348 GlnCysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGluCysLysCys 367
 Db 344 GATATTCTCTAAATAATGCTGGATGTTTATAGAGATGATAATGGTACTGAAGAANTGGAGATGT 403
 QY 368 LeuLeuAsnTyrLysGln---GluGlyAspLysCysValGluAsnProAsnProThrCys 386
 Db 404 TTATTAGGTACAAAAGAGTGAAGGTAAATACATGTGTAGAAAATAATATCTCTACTGT 463
 QY 387 AsnGluAsnAsnGlyCysAspAlaAspAlaThrCysThrGluGluAspSer 404
 Db 464 GATATCAACAATGCTGGATGTGATCCCACTGCTAGTTGTCAAAATCGGGAAC 517

RESULT 19
 BM160668
 LOCUS
 DEFINITION EST563191 PVBS Plasmodium yoelii yoelii cDNA clone PYCJU25 5' end, mRNA sequence.

ACCESSION BM160668
 VERSION BM160668.1 GI:17306349
 KEYWORDS EST.
 SOURCE Plasmodium yoelii yoelii.
 ORGANISM Plasmodium yoelii yoelii.
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 439)
 AUTHORS Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B., Fraser, C.M., and Carucci, D.J.
 TITLE Plasmodium yoelii EST project at TIGR
 JOURNAL Unpublished (2001)
 COMMENT Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADF.

FEATURES
 source
 1..439
 /organism="Plasmodium yoelii yoelii"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYCJU25"
 /clone_lib="pyBS"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BAUB/cbyJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of

5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybrizAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybrizAP vector and plasmid DNA isolated."

BASE COUNT 159 a 41 c 72 g 167 t
 ORIGIN

Alignment Scores:

Pred. No.: 1-32e-16 Length: 439
 Score: 243.50 Matches: 43
 Percent Similarity: 67.44% Conservative: 15
 Best Local Similarity: 50.00% Mismatches: 27
 Query Match: 10.70% Indels: 1
 DB: 13 Gaps: 1

US-10-057-531A-2 (1-431) x BM160668 (1-439)

QY 347 LysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGluCysLys 366
 Db 8 AGAGATATTCTCTAAATAATGCTGGATGTTTATAGAGATGATAATGCTACTGAAGAATGAGA 67
 QY 367 CysLeuLeuAsnTyrLysGln---GluGlyAspLysCysValGluAsnProAsnProThr 385
 Db 68 TGTTTATAGGTACAAAAGAGTGAAGGTAAATACATGTGTAGAAAATAATATCTCTACT 127
 QY 386 CysAsnGluAsnGlyCysAspAlaAspAlaThrCysThrGluGluAspSerGly 405
 Db 128 TGTGATATCAACAATGCTGGATGTCATCAACTGCTAGTTGTCAAAATCGGGAAGTACG 187
 QY 406 SerSerArgLysLysIleThrCysGluCysThrLysProAspSerTyrProLeuPheAsp 425
 Db 188 GAAATTTCCAAAAAATATATATACATGTATAAGAACCAACCCCTAATGCTATATTATGAA 247
 QY 426 GlyIlePheCysSerSer 431
 Db 248 GGTGATATTCTGTAGTTCT 265

RESULT 20
 BI094652
 LOCUS
 DEFINITION BI094652 Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P. resinosa Laccaria bicolor cDNA, mRNA sequence.

ACCESSION BI094652
 VERSION BI094652.1 GI:14516129
 KEYWORDS EST.
 SOURCE Laccaria bicolor.
 ORGANISM Laccaria bicolor.
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Agaricales; Tricholomataceae; Laccaria.
 REFERENCE 1 (bases 1 to 648)
 AUTHORS Podila, G.K., Brand, J.R. and Hynes, M.J.
 TITLE Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P. resinosa
 JOURNAL Unpublished (2001)
 COMMENT Contact: Dr. G.K. Podila
 Department of Biological Sciences
 University of Alabama
 Huntsville, AL 35899, USA
 Tel: 256 824 6263
 Fax: 256 824 6305
 Email: podila@emall.uah.edu.

FEATURES
 source
 1..648
 /organism="Laccaria bicolor"
 /db_xref="taxon:29883"
 /clone_lib="Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P. resinosa"
 /note="cDNA clones were selected using a combination of

Suppressive subtraction hybridization derived probes and a cDNA library of *L. bicolor*, prepared from pooled RNA of interaction time points ranging from 6-72 h. Differential expression of these cDNAs was confirmed through screening a membrane array of the cDNAs with exponential probes prepared from control RNA obtained from free living *L. bicolor* and pooled interaction RNA."

BASE COUNT 175 a 149 c 182 g 142 t
ORIGIN

Alignment Scores:
Pred. No.: 7.97e-16 Length: 648
Score: 238.50 Matches: 49
Percent Similarity: 80.33% Conservative: 0
Best Local Similarity: 80.33% Mismatches: 1
Query Match: 10.48% Indels: 12
DB: 13 Gaps: 1

US-10-057-531a-2 (1-431) x BI094652 (1-648)

Qy 1 MethHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
|||||
Db 338 ATGCACCATCATCATCTCTCTGGTCTGGTCCACCGGTTCTGGTATGAAGAA 397
|||||
Qy 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerPro----- 34
|||||
Db 398 ACCGCTGCTGCTAAATTCGAACGCCAGCACATGGACAGCCCATCTGCATTAGGTGAC 457
|||||
Qy 35 -----AspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGly 49
|||||
Db 458 ACTATAGTAATACCAAGATCTGGGTACCGACGACGACGACGACGATGCGCATTAGGTGAC 516
|||||

Qy 50 Ser 50
|||
Db 517 TCC 519

RESULT 21
BI094593
LOCUS
DEFINITION LBSSH00024 Preinfection stage symbiosis-regulated cDNAs from *L. bicolor* x *P. resinosa* Laccaria bicolor cDNA, mRNA sequence.
EST 20-JUN-2001

ACCESSION BI094593.1 GI:14516071

KEYWORDS

SOURCE

ORGANISM

Laccaria bicolor.
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Agaricales; Tricholomataceae; Laccaria.

REFERENCE 1 (bases 1 to 661)

AUTHORS Podila,G.K., Brand,J.R. and Hynes,M.J.

TITLE Preinfection stage symbiosis-regulated cDNAs from *L. bicolor* x *P. resinosa*

JOURNAL Unpublished (2001)

COMMENT

Contact: Dr. G.K. Podila

Department of Biological Sciences

University of Alabama

Huntsville, AL 35899, USA

Tel: 256 824 6263

Fax: 256 824 6305

Email: podilag@mail.uah.edu.

Location/Qualifiers

1. .661

/organism="Laccaria bicolor"

/db_xref="taxon:29883"

/clone_lib="Preinfection stage symbiosis-regulated cDNAs

from *L. bicolor* x *P. resinosa*"

/note="cDNA clones were selected using a combination of

Suppressive subtraction hybridization derived probes and

a cDNA library of *L. bicolor*, prepared from pooled RNA of

interaction time points ranging from 6-72 h. Differential

expression of these cDNAs was confirmed through screening

a membrane array of the cDNAs with exponential probes

prepared from control RNA obtained from free living *L.*

BASE COUNT 183 a 145 c 180 g 153 t
ORIGIN

Alignment Scores:

Pred. No.: 8.19e-16 Length: 661
Score: 238.50 Matches: 49
Percent Similarity: 80.33% Conservative: 0
Best Local Similarity: 80.33% Mismatches: 1
Query Match: 10.48% Indels: 12
DB: 13 Gaps: 1

US-10-057-531a-2 (1-431) x BI094593 (1-661)

Qy 1 MethHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
|||||
Db 412 ATGCACCATCATCATCTCTCTGGTCTGGTCCACCGGTTCTGGTATGAAGAA 471
|||||

Qy 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerPro----- 34
|||||

Db 472 ACCGCTGCTGCTAAATTCGAACGCCAGCACATGGACAGCCCATCTGCATTAGGTGAC 531
|||||

Qy 35 -----AspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGly 49
|||||

Db 532 ACTATAGTAATACCAAGATCTGGGTACCGACGACGACGACGACGATGCGCATTAGGTGAC 590
|||||

Qy 50 Ser 50
|||

Db 591 TCC 593

RESULT 22

BI094577

LOCUS

DEFINITION LBSSH00008 Preinfection stage symbiosis-regulated cDNAs from *L. bicolor* x *P. resinosa* Laccaria bicolor cDNA, mRNA sequence.

EST 20-JUN-2001

ACCESSION BI094577.1 GI:14516055

KEYWORDS

SOURCE

ORGANISM

Laccaria bicolor.
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Agaricales; Tricholomataceae; Laccaria.

REFERENCE 1 (bases 1 to 686)

AUTHORS Podila,G.K., Brand,J.R. and Hynes,M.J.

TITLE Preinfection stage symbiosis-regulated cDNAs from *L. bicolor* x *P. resinosa*

JOURNAL Unpublished (2001)

COMMENT

Contact: Dr. G.K. Podila

Department of Biological Sciences

University of Alabama

Huntsville, AL 35899, USA

Tel: 256 824 6263

Fax: 256 824 6305

Email: podilag@mail.uah.edu.

Location/Qualifiers

1. .686

/organism="Laccaria bicolor"

/db_xref="taxon:29883"

/clone_lib="Preinfection stage symbiosis-regulated cDNAs

from *L. bicolor* x *P. resinosa*"

/note="cDNA clones were selected using a combination of

Suppressive subtraction hybridization derived probes and

a cDNA library of *L. bicolor*, prepared from pooled RNA of

interaction time points ranging from 6-72 h. Differential

expression of these cDNAs was confirmed through screening

a membrane array of the cDNAs with exponential probes

prepared from control RNA obtained from free living *L.*

bicolor and pooled interaction RNA."

BASE COUNT 194 a 154 c 179 g 157 t

ORIGIN

Alignment Scores:

Pred. No.: 8.61e-16 Length: 686

Score: 238.50 Matches: 49
 Percent Similarity: 80.33% Conservative: 0
 Best Local Similarity: 80.33% Mismatches: 12
 Query Match: 10.48% Indels: 1
 DB: 13 Gaps: 1

US-10-057-531A-2 (1-431) x BI094577 (1-686)

QY 1 MethHisHisHisHisHisSerGlyLeuValProArgGlySerGlyMetLysGlu 20

Db 406 ATGCACCATCATCATCATCTCTCTGGTCGTGCCCGCGGTCTGGTATGAAGAA 465

QY 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerPro----- 34

Db 466 ACGCTGCTGCTAAATTCGAACCCAGCACATGGACAGCCACATCTGCATTTAGGTGAC 525

QY 35 -----AspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGly 49

Db 526 ACTATAGAATACCAAGATCTGGGTACCGAGCAGCAGCAAGGCCATGGC-GATATCGGA 584

QY 50 Ser 50

Db 585 TCC 587

RESULT 23

BI094637

LOCUS

DEFINITION

Accession

Version

Keywords

Source

Organism

Reference

Authors

Title

Journal

Comment

Features

Source

Location/Qualifiers

1. .415

/organism="Laccaria bicolor"

/db_xref="taxon:29883"

/clone_lib="Preinfection stage symbiosis-regulated cDNAs

from L. bicolor x P. resinosa"

/note="cDNA clones were selected using a combination of

suppressive subtraction hybridization derived probes and

a cDNA library of L. bicolor, prepared from pooled RNA of

interaction time points ranging from 6-72 h. Differential

expression of these cDNAs was confirmed through screening

a membrane array of the cDNAs with exponential probes

prepared from control RNA obtained from free living L.

bicolor and pooled interaction RNA."

BASE COUNT 125 a 95 c 108 g 86 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 1.2e-15

Score: 234.50

Percent Similarity: 78.69%

Best Local Similarity: 78.69%

Query Match: 10.31%

DB: 13

Matches: 48

Conservative: 2

Mismatches: 12

Indels: 1

Gaps: 1

US-10-057-531A-2 (1-431) x BI094637 (1-415)

QY 1 MethHisHisHisHisHisSerGlyLeuValProArgGlySerGlyMetLysGlu 20

Db 158 ATGCACCATCATCATCATCTCTCTGGTCGTGCCCGCGGTCTGGTATGAAGAA 217

QY 21 ThrAlaAlaLysPheGluArgGlnHisMetAspSerPro----- 34

Db 218 ACGCTGCTGCTAAATTCGAACCCAGCACATGGACAGCCACATCTGCATTTAGGTGAC 277

QY 35 -----AspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGly 49

Db 278 ACTATAGAATACCAAGATCTGGGTACCGAGCAGCAGCAAGGCCATGGC-GATATCGGA 336

QY 50 Ser 50

Db 337 TCC 339

RESULT 24

BF298872

LOCUS

DEFINITION

Accession

Version

Keywords

Source

Organism

Reference

Authors

Title

Journal

Comment

Features

Source

Location/Qualifiers

1. .660

/organism="Plasmodium berghei"

/strain="ANKA clone 15cyl1 (clone of the ANKA 8417 HP clone

)"

/db_xref="taxon:5821"

/clone_lib="PB cDNA #20, Charles Yowell and Jane Carlton"

/dev_stage="asynchronous blood stage"

/lab_host="Swiss white mice"

/note="Vector: pBluescript II vector DNA, excised from

Lambda ZAP II.; Site 1: EcoRI; Site 2: XhoI; Total RNA was

extracted from asynchronous blood stage forms of the

Swiss white mice. Contaminating host white cells had

previously been removed using a novel biomagnetic bead

protocol (J. Carlton et al., manuscript in preparation).

PolyA+ RNA was extracted and reverse transcribed using an

oligo dT-XhoI primer. Second strand cDNA was prepared

using RNase H and DNA polymerase I. EcoRI adaptors were

ligated to the cDNA, and it was digested with XhoI.

Fragments were size selected, and those between 1-5 kb

ligated into EcoRI /XhoI digested vector."

BASE COUNT 307 a 71 c 99 g 181 t 2 others

ORIGIN

Alignment Scores:

Pred. No.: 2.25e-15

Score: 234.50

Percent Similarity: 51.65%

Best Local Similarity: 24.79%

Query Match: 10.31%

DB: 12

Matches: 660

Conservative: 65

Mismatches: 85

Indels: 32

Gaps: 5

US-10-057-531a-2 (1-431) x BF298872 (1-660)

Qy 111 ArgLeuLysLysArgLysTyrPheLeuAspValLeuGluSerAspLeuMetClnPhelLys 130
 Db 1 AAATACAAAGAAAGAAATGTTCTTCTGAAGTATTAAATCATGAATGGATCTATTCAAA 60
 Qy 131 HisLeuSerAsnGluTyrLysLeuGluAspSerPheLysLeuLeuAsnSerGluGln 150
 Db 61 GATTTAAGTACCACAAATATGTTATTAGAACCCATACCAATATTAGATGATATAA 120
 Qy 151 LysAsnThrLeuLysSerTyrLysTyrLysLysGluSerValGluAsnAspLys 170
 Db 121 AAAGACAAACAATAGTAAATTTAAATATGCTGCTAAAGGTGTAATCAAGATATAGAA 180
 Qy 171 PheAlaGlnGluLysSerTyrTyrGluLysValLeuAlaLysTyrLysAspLeu 190
 Db 181 ACAACTGCTGACGGAATTAATCTTTAACAAATGATTGAATTATACAAATCAATTA 240
 Qy 191 GluSerLysLysValLysLysGluLysGluLysPheProSerSerProThr 210
 Db 241 GCTGAGTAAAGAAATGATGCCATAGAA----- 273
 Qy 211 ThrProSerProAlaLysThrAspGlnLysLysGluSerLysPheLeuProPhe 230
 Db 274 -----GCTGCTACTACCGATAAGATGAAAGAAAGAAATATGTTCCAATC 318
 Qy 231 LeuThrAsnLysGluThrLysAsnLysValLysLysLysLysLysLysLysLys 250
 Db 319 TTGAGATCTTAAGGATTAATGAAACCATATTAGTCAATCAGAGAATATATAGAA 378
 Qy 251 AsnLeuLysAlaLysLysAsnValGluLysAspGluAlaHisValLysLys 270
 Db 379 GTATTACAAATAGACTTGATAGTTATATAAATGAAAGACGTAATNTCAATTTAAGC 438
 Qy 271 ThrLysLeuSerAspLeuLysAlaLysLysLysLysLysLysLysLysLysLys 290
 Db 439 AAAATTTAGAACATACATAAAATGACGAAACCTTGAAATTTTC----- 486
 Qy 291 AspPheGluAlaLysLysLysLysLysLysLysLysLysLysLysLysLys 307
 Db 487 -----GTAGAAATGCAGAAATATAAATACATACATAGCCCTCAATG 525
 Qy 308 ---LeuGlyLysLeuLysSerThrGlyLeuVal---GlnAsnPheProAsnThrLys 325
 Db 526 GCTTTAAATAACTTAATAAATCTGTTTGTGAGGAGGCGAATCGAAATAATATTA 585
 Qy 326 SerLysLeuLysGluLysPheGlnAspMetLeuAsnLysLysLysLysLysLys 344
 Db 586 GCAAAATGCTTAATATGATAGTGGATTTATTAAATAGATAGACCTAAGCATGTTTG 645
 Qy 344 sVal 345
 Db 646 TTTA 649
 RESULT 25
 BI094651
 LOCUS
 DEFINITION L85SH00094 Preinfection stage symbiosis-regulated cDNAs from L.
 bicolor x P. resinosa Laccaria bicolor cDNA, mRNA sequence.
 ACCESSION BI094651
 VERSION BI094651.1 GI:14516128
 KEYWORDS EST.
 SOURCE Laccaria bicolor.
 ORGANISM Laccaria bicolor
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 Agaricales; Tricholomataceae; Laccaria.
 REFERENCE 1 (bases 1 to 405)
 AUTHORS Podila,G.K., Brand,J.R. and Hynes,M.J.
 TITLE Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P.
 resinosa
 JOURNAL Unpublished (2001)
 COMMENT Contact: Dr. G.K. Podila

Department of Biological Sciences
 University of Alabama
 Huntsville, AL 35899, USA
 Tel: 256 824 6263
 Fax: 256 824 6305
 Email: podilag@mail.uah.edu.

FEATURES
 source
 1. 405
 /organism="Laccaria bicolor"
 /db_xref="taxon:29883"
 /clone_lib="Preinfection stage symbiosis-regulated cDNAs
 from L. bicolor x P. resinosa"
 /note="cDNA clones were selected using a combination of
 suppressive subtraction hybridization derived probes and
 a cDNA library of L. bicolor, prepared from pooled RNA of
 interaction time points ranging from 6-72 h. Differential
 expression of these cDNAs was confirmed through screening
 a membrane array of the cDNAs with exponential probes
 prepared from control RNA obtained from free living L.
 bicolor and pooled interaction RNA."

BASE COUNT 116 a 92 c 112 g 85 t
 ORIGIN

Alignment Scores:
 Pred. NO.: 1.5e-15 Length: 405
 Score: 233.50 Matches: 56
 Percent Similarity: 53.51% Conservative: 5
 Best Local Similarity: 49.12% Mismatches: 12
 Query Match: 10.26% Indels: 42
 DB: 13 Gaps: 3

US-10-057-531a-2 (1-431) x BI094651 (1-405)

Qy 1 MethHisHisHisHisHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
 Db 126 ATGCACCATCATCATCTCTCTGTCGGTCCACCGCGTGTGGTATGAAGAA 185
 Qy 21 ThrAlaLysLysPheGluArgGlnHisMetAspSerPro----- 34
 Db 186 ACCGCTGCTGCTAAATTCGACGCCAGCACATGGACGCCAGATCTGCTTAGTGAC 245
 Qy 35 -----AspLeuGlyThrAspAspAspLysAlaMetAlaAspLysGly 49
 Db 246 CCTATAGATACCAAGATCTGGTACCGACGACGACGACGATGCGC-GATATCGGA 304
 Qy 50 SerLysGluGlyArgGlyThrMetAlaLysSerValThrMetAspAsnLysLeuSerGly 69
 Db 305 TCC----- 307
 Qy 70 PheGluAsnGluTyrAspValLysTyrLeuLys-----Pro 81
 Db 308 -----GAATTCGCGCGCGCTCGATACGGCTCGAGAGAGACGACAGAGGGTCCC 358
 Qy 82 LeuAlaGlyValTyrArgSerLeuLysLysGlnLysGluLys 95
 Db 359 CTTGCGGTGACATACGAAAAA----- 400

RESULT 26
 BI094651
 LOCUS
 DEFINITION EST562555 PyBS Plasmodium yoelii yoelii cDNA clone P1CJM10 5' end,
 mRNA sequence.
 ACCESSION BI094651
 VERSION BI094651.1 GI:17305713
 KEYWORDS EST.
 SOURCE Plasmodium yoelii yoelii.
 ORGANISM Plasmodium yoelii yoelii
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 725)
 AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
 Fraser,C.M. and Carucci,D.J.
 TITLE Plasmodium yoelii EST project at TIGR
 JOURNAL Unpublished (2001)

collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over PAD-GAL4; AL20-23% parasitemia, blood was

ACCESSION

BM164097 782 bp mRNA linear EST 04-DEC-2001
EST366620 yB5 Plasmodium yoelii cDNA clone PYCLP63 5' end,
mRNA sequence.
BM164097

VERSION BM164097.1 GI:17309778
EST.
KEYWORDS
SOURCE

ORGANISM
Plasmodium yoelii yoelii.

REFERENCE
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

AUTHORS
1 (bases 1 to 782)

TITLE
Fraser, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B.,
Plasmodium yoelii EST project at TIGR

JOURNAL
Unpublished (2001)

COMMENT
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org

For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

FEATURES

source

1..782

/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone_lib="PYCLP63"
/clone_lib="PyBS"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cByJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidinium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (pAD-GAL4) was
excised from the HybridAP vector and plasmid DNA
isolated."

BASE COUNT 334 a 119 c 141 g 188 t

Alignment Scores:

Pred. No.: 8.34e-12 Length: 782
Score: 203.00 Matches: 54
Percent Similarity: 51.79% Conservative: 33
Best Local Similarity: 32.14% Mismatches: 65
Query Match: 8.92% Indels: 16
DB: 13 Gaps: 4

US-10-057-531A-2 (1-431) x BM164097 (1-782)

Qy 10 GlyLeuValProArgGlySerGlyMetLysGluThrAlaAlaAlaLysPheGluArgGln 29
Db 326 GGTGAGTACCA---GGATGACGACCGATACACGGGTAGCTGGA----- 367
Qy 30 HisMetAspSerProAspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGly 49
Db 368 -----ACCATGTTGATGATAAGACCGATGATATATCAATTCGAAGT---CGT 418
Qy 50 SerIleGluGlyArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSerGly 69
Db 419 CAATCCGAA-----GATGCACCCAGAAAAAGATATCTTCCCAA 457
Qy 70 PheGluAsnGluTyAspValIleTyLeuLysProLeuAlaGlyValTyArgSerLeu 89

Db 458 TTTCAAAATGAAAGTTTGTATGTATACACAAAAAGTTGGGTAGTACATATAAAATCATTA 517
Qy 90 LysLysGlnIleGluLysAsnIlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsn 109
Db 518 AAGAACAACATGTTAAGAGAAATTTTCAACAATTAAGAAGACATGACAAATGGATTAAAT 577
Qy 110 SerArgLeuLysLysArgLysTyThrPheLeuAspValLeuGluSerAspLeuMetGlnPhe 129
Db 578 AATAATCACAAAAAAGAAATGATTTCTTGAAGTATTAAAGCCATGATTAATTC 637
Qy 130 LysHisIleSerSerAsnGluTyThrIleIleGluAspSerPheLysLeuLeuAsnSerGlu 149
Db 638 AAAGATTTAAGTACCACAAATATGTTATTAGAAATCCATATCAATATTAGATTAATGAT 697
Qy 150 GlnLysAsnThrLeuLysSerTyThrLysGluSerValGluAsnAspIle 169
Db 698 AAAAAGACAACCACTAAGTAACTTAAGTAAAGTATTAAGGTATTAAGTAAAGATATA 757
Qy 170 LysPheAlaGlnGluGlyIleSer 177
Db 758 GAAACAACACTACTGACGGAATTAAT 781

RESULT 32

BM167649

LOCUS

DEFINITION EST570172 PYBS Plasmodium yoelii yoelii cDNA clone PYCO038 5' end,
mRNA sequence.

ACCESSION BM167649

VERSION BM167649.1

KEYWORDS EST.

SOURCE Plasmodium yoelii yoelii.

ORGANISM Plasmodium yoelii yoelii

REFERENCE 1 (bases 1 to 329)

AUTHORS

Fraser, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B.,
Fraser, C.M. and Carucci, D.J.

TITLE Plasmodium yoelii EST project at TIGR

JOURNAL Unpublished (2001)

COMMENT Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org

For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html

Seq primer: ADF.

Location/Qualifiers

1..329

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone_lib="PYCO038"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cByJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidinium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (pAD-GAL4) was
excised from the HybridAP vector and ligated to
HybridAP arms directionally using EcoRI-XhoI cleaved arms."

After packaging, the phagemid vector (pAD-GAL4) was excised from the HybridAP vector and plasmid DNA isolated."

```

BASE COUNT      128 a   40 c   67 g   94 t
ORIGIN

Alignment Scores:
Pred. No.:      1.96e-11      Length:      329
Score:          195.00      Matches:      43
Percent Similarity: 62.11%      Conservative: 16
Best Local Similarity: 45.26%      Mismatches:  32
Query Match:      8.57%      Indels:      4
DB:              13      Gaps:      4

US-10-057-531A-2 (1-431) x BM167649 (1-329)
QY 308 LeuGlyLysLeuLeuSerThrGlyLeuVal---GlnAsnPheProAsnThrIleLeuSer 326
Db 43 TTAACAACCTTAATAATCTGGTTAGTAGGAGAGGTGAATCAAGAAGAAATATTAGCA 102
QY 327 LysLeuIleGluGlyLysPheGlnAspMetLeuAsnIleSer---GlnHisGlnCysVal 345
Db 103 AATGCTTAACATGATGATGATGATTTATTACGTGTAGACCCCTAAACATGTTATCGGTT 162
QY 346 ---LysLysGlnCysProGluAsnSerGlyCysPheArgHisLeuAspGluArgGluGlu 364
Db 163 GATACAAGAGATATCTCTACAAATCTGGATGTTTATAGAGATGATAATGCTACTGAAGAA 222
QY 365 CysLysCysLeuLeuAsnThrLysGln---GluGlyAspLysCysValGluAsnProAsn 393
Db 223 TGGAGATGTTTATTACGTTACAAACAAGGTGAAGGTAAATACATGTGTAGAAATAATAAT 282
QY 384 ProThrCysAsnGluAsnAsnGlyLysCysAspAlaAspAlaThr 398
Db 283 CCTACTGTGATATCAACAATGCTGGATGTGATCCAACTGCTAGT 327

RESULT 33
BI094571
LOCUS
DEFINITION
LbSSH00002 Preinfection stage symbiosis-regulated cDNAs from L.
bicolor x P. resinosa Laccaria bicolor cDNA, mRNA sequence.
ACCESSION
BI094571
VERSION
BI094571.1 GI:14516049
SOURCE
Laccaria bicolor.
ORGANISM
Laccaria bicolor
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Agaricales; Tricholomataceae; Laccaria.
REFERENCE
Podila,G.K., Brand,J.R. and Hynes,M.J.
Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P.
resinosa
JOURNAL
Unpublished (2001)
COMMENT
Contact: Dr. G.K. Podila
Department of Biological Sciences
University of Alabama
Huntsville, AL 35899, USA
Tel: 256 824 6263
Fax: 256 824 6305
Email: podilag@mail.uah.edu.
Location/Qualifiers
1. .624
/organism="Laccaria bicolor"
/db_xref="taxon:29883"
/clone_lib="Preinfection stage symbiosis-regulated cDNAs
from L. bicolor x P. resinosa"
/note="cDNA clones were selected using a combination of
Suppressive subtraction hybridization derived probes and
a cDNA library of L. bicolor, prepared from pooled RNA of
interaction time points ranging from 6-72 h. Differential
expression of these cDNAs was confirmed through screening
a membrane array of the cDNAs with exponential probes
prepared from control RNA obtained from free living L.

BASE COUNT      192 a   124 c   156 g   151 t   1 others
ORIGIN

Alignment Scores:
Pred. No.:      8.79e-11      Length:      624
Score:          192.50      Matches:      59
Percent Similarity: 58.14%      Conservative: 16
Best Local Similarity: 45.74%      Mismatches:  37
Query Match:      8.46%      Indels:      18
DB:              13      Gaps:      3

US-10-057-531A-2 (1-431) x BI094571 (1-624)
QY 3 HisHisHisHisSerSerGlyLeuVal-ProArgLysSerGlyMetLysGluThrAl 22
Db 228 AATCATCATCATCATCTCTCTGCTGCTGCCACCGCGTTCGTGTATGAAGAAACCGC 287
QY 22 aAlaAlaLysPheGluArg-GlnHisMetAspSerPro----- 34
Db 288 TGCTGCTAAATTCGAACGCCACACATGGACAGCCACAGATCTGCATTTAGGTGACACTA 347
QY 35 -----AspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGlySerI 51
Db 348 TAGAATACCAAGATCTGGGTACCGACGACGACGACGACGATGCG- 405
QY 51 leGluGlyArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSerGlyPheG 71
Db 406 -----GNAATTCGGCGCGCTCGATACGCTGGAGAGACGACGAGGCGGGAAGCAG 460
QY 71 luAsnGluTyArgPheValIleTyLeuLys---ProLeuAlaGlyValTyArgSerLeuL 90
Db 461 CAAGCAGGGTAAACGTGATTTGGATGAACCTTGAGGCACGATGATCTATGAGGATTTGGA 520
QY 90 yLysGlnIleGluLysAsnIlePheThrPheAsnLeu-AsnLeuAsnAspIleLeuAsn 109
Db 521 TGAGCTTGACTGAGCGGGGATTTATCGTTATATGATATGATCTAGTAGGAAGATAATAA 580

RESULT 34
BI094571
LOCUS
DEFINITION
EST571502 PyBS Plasmodium yoelii yoelii cDNA clone PYCPK24 5' end,
mRNA sequence.
ACCESSION
BI094571
VERSION
BI094571.1 GI:17302211
KEYWORDS
Plasmodium yoelii yoelii.
SOURCE
Plasmodium yoelii yoelii.
ORGANISM
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE
1 (bases 1 to 780)
Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
Fraser,C.M. and Carucci,D.J.
Plasmodium yoelii EST project at TIGR
JOURNAL
Unpublished (2001)
COMMENT
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.
Location/Qualifiers
1. .780
/organism="Plasmodium yoelii yoelii"
/strain="17XL"

FEATURES
source

```


QY 1 MethHisHsHisHsHisSerSerGlyLeuValProArg-GlySerGlyMetLysG1 20
 Db 193 ATGCACCATCATCATCATCTCTCTGGTGGCCACGCGGGTTCTGCTATGAAGA 252
 QY 20 uThr-AlaAlaAlaLysPheGluArgGln-HisMetAspSer----- 33
 Db 253 AACCGGCTGCTGCTAAATTCGAACGCCAACACATGTCGACAGCCGCCAGATCTGCATTTAGG 312
 QY 34 -----Pro-AspLeuGlyThrAspAspAspAspLysAlaMetAlaAspI1 48
 Db 313 TGACACTATAGATACCAAGATCTGGGTACCGACGACGACAGCCCATGCC-GATAT 371
 QY 48 eGlySer 50
 Db 372 CGGATCC 378

RESULT 36
 BI094595 608 bp mRNA linear EST 20-JUN-2001
 LOCUS LbSSH00026 Preinfection stage symbiosis-regulated cDNAs from L.
 DEFINITION bicolor x P. resinosa Laccaria bicolor cDNA, mRNA sequence.
 ACCESSION BI094595
 VERSION BI094595.1 GI:14516073
 KEYWORDS EST.
 SOURCE Laccaria bicolor.
 ORGANISM Laccaria bicolor.
 Agaricales; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 Eukaryota; Tricholomataceae; Laccaria.
 REFERENCE 1 (bases 1 to 608)
 AUTHORS Podila, G.K., Brand, J.R. and Hynes, M.J.
 TITLE Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P.
 resinosa
 JOURNAL Unpublished (2001)
 COMMENT Contact: Dr. G.K. Podila
 Department of Biological Sciences
 University of Alabama
 Huntsville, AL 35899, USA
 Tel: 256 824 6263
 Fax: 256 824 6305
 Email: podila@emh.uah.edu.

FEATURES
 source
 1..608
 /organism="Laccaria bicolor"
 /db_xref="taxon:29883"
 /clone_lib="Preinfection stage symbiosis-regulated cDNAs
 from L. bicolor x P. resinosa"
 /note="cDNA clones were selected using a combination of
 suppressive subtraction hybridization derived probes and
 a cDNA library of L. bicolor, prepared from pooled RNA of
 infection time points ranging from 6-72 h. Differential
 expression of these cDNAs was confirmed through screening
 a membrane array of the cDNAs with exponential probes
 prepared from control RNA obtained from free living L.
 bicolor and pooled interaction RNA."
 BASE COUNT 172 a 141 c 167 g 128 t
 ORIGIN

Alignment Scores:
 Pred. No.: 2.66e-10 Length: 608
 Score: 189.00 Matches: 47
 Percent Similarity: 77.05% Conservativeness: 0
 Best Local Similarity: 77.05% Mismatches: 3
 Query Match: 8.26% Indels: 14
 DB: 13 Gaps: 1

US-10-057-531A-2 (1-431) x BI094595 (1-608)

QY 1 MethHisHsHisHsHisSerSerGlyLeuValProArgGlySerGlyMetLysGlu 20
 Db 341 ATGCACCATCATCATCATCTCTCTGGTGGCCACGCGGGTTCTGGTATGAA-GAA 399
 QY 21 ThrAlaAlaAlaLysPheGluArgGlnHisMetAspSerProAspLeu----- 36
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 400 ACCGCTGCTGCTAAATTCGAACGCCACGACATGTCGACAGCCAGATCT-GCATTTAGGTGA 458
 QY 37 -----GlyThrAspAspAspLysAlaMetAlaAspIlegly 49
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 459 CACTATAGATACCAAGTCTGGGTACCGACGACGACAGCCATGGC-GATATCGGA 517
 QY 50 Ser 50
 |||
 Db 518 TCC 520

RESULT 37
 BM166573 652 bp mRNA linear EST 04-DEC-2001
 LOCUS EST569096 PyBS Plasmodium yoelii yoelii cDNA clone pYCNT91 5' end.
 DEFINITION mRNA sequence.
 ACCESSION BM166573
 VERSION BM166573.1 GI:17299805
 KEYWORDS EST.
 SOURCE Plasmodium yoelii yoelii.
 ORGANISM Plasmodium yoelii yoelii.
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 652)
 AUTHORS Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B.,
 Fraser, C.M. and Carucci, D.J.
 TITLE Plasmodium yoelii EST project at TIGR
 JOURNAL Unpublished (2001)
 COMMENT Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/Index.html
 Seq primer: ADF.

FEATURES
 source
 1..652
 Location/Qualifiers
 /organism="Plasmodium yoelii yoelii"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone_lib="PyBS"
 /lab_host="E. coli XL-1 Blue"
 /note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
 collected from BALB/cByJ mice infected with Py17XL
 parasites, and leukocytes removed by passage over
 microcrystalline cellulose columns. Total RNA was
 isolated using the guanidinium isothiocyanate method, and
 mRNA isolated using oligo(dT)-cellulose chromatography.
 First strand cDNA synthesis was completed using a 50-base
 primer and reverse transcriptase in the presence of
 5-methyl dCTP. After second strand synthesis, uneven
 termini were treated with pfu DNA polymerase, and EcoRI
 adaptors ligated to the blunt ends. The sample was cleaved
 with XhoI and separated on a Sephacryl S-500 column.
 Size-fractionated cDNA was precipitated and ligated to
 HybriZAP arms directionally using EcoRI-XhoI cleaved arms.
 After packaging, the phagemid vector (pAD-GAL4) was
 excised from the HybriZAP vector and plasmid DNA
 isolated."
 BASE COUNT 290 a 97 c 117 g 148 t
 ORIGIN

Alignment Scores:
 Pred. No.: 8.05e-10 Length: 652
 Score: 184.00 Matches: 51
 Percent Similarity: 50.31% Conservativeness: 29
 Best Local Similarity: 32.08% Mismatches: 63
 Query Match: 8.09% Indels: 16
 DB: 13 Gaps: 4

BASE COUNT	218 a	130 c	160 g	129 t
US-10-057-531A-2 (1-431) x BM166573 (1-652)				
QY	10	GlyLeuValProArgSerGlyMetLysGluThrAlaAlaLysPheGluArgGln	29	
Db	176	GGTGCAGTACCA---GGATCAGAACCGGATACACGGGTAGCTGGA-----	217	
QY	30	HisMetAspSerProAspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGly	49	
Db	218	-----ACGAGGTGTGATGATGAATGACGACGATGATATATCAAAATGCAAGT---	268	
QY	50	SerIleGluGlyArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSerGly	69	
Db	269	CATCCGAA-----GATGCACACAGAAAGATATCTTCTCCCAA	307	
QY	70	PheGluAsnGluThrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeu	89	
Db	308	TTTACAAATCAAAAGTGTGTATATACACAAAGGTTGGGTAGTACATATAAATCATTA	367	
QY	90	LysLysGlnIleLysAsnIlePheThrPheAsnLeuAsnLeuAspIleLeuAsn	109	
Db	368	AGAAACACATGTTAAGACAAATTTCAACAATTAAGAGACATGACAAATGGATTAAT	427	
QY	110	SerArgLeuLysLysArgLysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPhe	129	
Db	428	ATAAATATCAAAAGAAATGATTTCCTTGAAGTATTAAAGCCATGAATTAGATTATTC	487	
QY	130	LysHisIleSerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGlu	149	
Db	488	AAAGATTAAATACCAACAATATGTTATAGAAATCCATATCAATATTAGATAATGAT	547	
QY	150	GlnLysAsnThrLeuLysSerTyrLysTyrIleLysGluSerValGluAsnAsp	168	
Db	548	AAAAAGACAAACAAATAGTAAACTTAAATATGCTACTAAAGGTAAAAATGAAGAT	604	
RESULT 38				
LOCUS	BI094585			
DEFINITION	LbSSH00016 Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P. resinosa Laccaria bicolor cDNA, mRNA sequence.			
ACCESSION	BI094585			
KEYWORDS	EST.			
SOURCE	BI094585.1 GI:14516063			
ORGANISM	Laccaria bicolor.			
REFERENCE	Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Agaricales; Tricholomataceae; Laccaria.			
AUTHORS	Podila, G.K., Brand, J.R. and Hynes, M.J.			
TITLE	Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P. resinosa			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Dr. G.K. Podila Department of Biological Sciences University of Alabama Huntsville, AL 35899, USA Tel: 256 824 6263 Fax: 256 824 6305 Email: podilag@mail.uah.edu.			
FEATURES				
source	Location/Qualifiers			
	1..367			
	/organism="Laccaria bicolor"			
	/db_xref="taxon:29883"			
	/clone_lib="Preinfection stage symbiosis-regulated cDNAs from L. bicolor x P. resinosa"			
	/note="cDNA clones were selected using a combination of suppressive subtraction hybridization derived probes and a cDNA library of L. bicolor, prepared from pooled RNA of interaction time points ranging from 6-72 h. Differential expression of these cDNAs was confirmed through screening a membrane array of the cDNAs with exponential probes prepared from control RNA obtained from free living L. bicolor and pooled interaction RNA."			
US-10-057-531A-2 (1-431) x BI094585 (1-637)				
QY	3	HisHisHisHisSerSerGlyLeuValProArg-GlySerGlyMetLysGluThrAl	22	
Db	111	AATCATCATCATCATCTCTCTGGTCTGGGGCCCAACGGTTCTGTGTTATTAACAAACCGC	170	
QY	22	aaLaAlaLysPheGluArgGlnHisMetAspSerPro	34	
Db	171	TGCTGCTAAATTCGAACGCCACGACATGACAGCCCATTTGCATTTAGTAACTAT	230	
QY	35	-----AspLeuGlyThrAspAspAspLysAlaMetAlaAspIleGlySer	50	
Db	231	AGAATACCAAGATCTGGTACCGACGACAGCAAGCCATGCGATATCGGATTCGATCC	288	
RESULT 39				
LOCUS	BI094585			
DEFINITION	PfESToab05a09.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE-PROTEIN-1 ; mRNA sequence.			
ACCESSION	BI094585			
VERSION	BI094585.1			
KEYWORDS	EST.			
SOURCE	BI094585.1 GI:21254317			
ORGANISM	malaria parasite P. falciparum.			
REFERENCE	1 (bases 1 to 362)			
AUTHORS	Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I., Tsagareishvili, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A., Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D., Waterston, R., Wilson, R. and Sibley, D.			
TITLE	WashU Plasmodium EST Project			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: L. David Sibley WashU Plasmodium EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University Seq primer: -400P from Gibco. Location/Qualifiers 1..362 /organism="Plasmodium falciparum" /db_xref="taxon:5833" /clone_lib="Plasmodium falciparum 3D7 asexual cDNA" /lab_host="DH10B (GeneHog, Invitrogen, Inc.)" /note="vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage zoonin(0.1%) - lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-Tract mRNA isolation system (Promega, WI) using streptavidin Magnosphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg)			

into EcoRI and XhoI sites of 1 zapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the Exassist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

BASE COUNT 153 a 32 c 33 g 144 t

ORIGIN

Alignment Scores:

Pred. No.: 3,47e-08 Length: 362
Score: 166.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.30% Indels: 0
DB: 14 Gaps: 0

US-10-057-531A-2 (1-431) x B0451205 (1-362)

QY 402 GluAspSerGlySerArgLysLysIleThrCysGluCysThrLysProAspSerTyr 421

Db 2 GAGATTCAGGTAGCAGCAGAAAGAAATCACATGTGATCTACTAAACCTGATTCTTAT 61

QY 422 ProLeuPheAspGlyLePheCysSerSer 431

Db 62 CCACCTTTTCGATGATGTTTCTGCAGTTC 91

RESULT 40

BM167595

LOCUS

DEFINITION

BM167595 686 bp mRNA linear EST 04-DEC-2001

EST570118 PyBS Plasmodium yoelii yoelii cDNA clone PYCON35 5' end,

mRNA sequence.

BM167595

BM167595.1 GI:17300827

EST.

Plasmodium yoelii yoelii.

Plasmodium yoelii yoelii

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

1 (bases 1 to 686)

Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,

Fraser,C.M. and Carucci,D.J.

Plasmodium yoelii EST project at TIGR

Unpublished (2001)

Contact: Jane Carlton

Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319

Fax: 301-838-0208

Email: carlton@tigr.org

For clone info, please contact the Malaria Research and Reference

Reagent Resource Center, ATCC

http://www.malaria.mr4.org/mr4pages/index.html

Seq primer: ADF.

Location/Qualifiers

1. .686

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCON35"

/dev_stage="PyBS"

/lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was

collected from BALB/cByJ mice infected with Py17XL

parasites, and leukocytes removed by passage over

microcrystalline cellulose columns. Total RNA was

isolated using the guanidinium isothiocyanate method, and

mRNA isolated using oligo(dT)-cellulose chromatography.

First strand cDNA synthesis was completed using a 50-base

primer and reverse transcriptase in the presence of

5-methyl dCTP. After second strand synthesis, uneven

termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybrizAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybrizAP vector and plasmid DNA isolated."

BASE COUNT 283 a 108 c 130 g 165 t

ORIGIN

Alignment Scores:

Pred. No.: 1.37e-07 Length: 686
Score: 164.00 Matches: 47
Percent Similarity: 50.00% Conservative: 23
Best Local Similarity: 33.57% Mismatches: 54
Query Match: 7.21% Indels: 16
DB: 13 Gaps: 4

US-10-057-531A-2 (1-431) x BM167595 (1-686)

QY 10 GlyLeuValProArgGlySerGlyMetLysGluThrAlaAlaLysPheGluArgGln 29

Db 315 GGTGCGAGTACCA---GGATCAGGAACCGATACACGGGTAGCTGGA----- 356

QY 30 HisMetAspSerProAspLeuGlyThrAspAspLysLysAlaMetAlaAspIleGly 49

Db 357 -----AGCAGTGTGATGATGAAGACGATGATATATCAAAATGCAAGT---GGT 407

QY 50 SerIleGluGlyArgGlyThrMetAlaIleSerValThrMetAspAsnIleLeuSerGly 69

Db 408 CAATCCGAA-----GATGCACCCAGAAAAGATATTTCTTCGGAA 446

QY 70 PheGluAsnGluTyrAspValIleTyrLeuLysProLeuAlaGlyValTyrArgSerLeu 89

Db 447 TTTACAANTGAAAGTTGTTGATGATATACAAAAGGTTGGGTAGTACATATAAATCATTA 506

QY 90 LysLysGlnIleGluLysAsnIlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsn 109

Db 507 AAGAAACACATGTTAAGAGAATTTTCAACAATTAAGACACATGACAAATGGATTAAAT 566

QY 110 SerArgLeuLysLysArgLysTyrPheLeuAspValLeuGluSerAspLeuMetGlnPhe 129

Db 567 AATAAATCACAATAAGAAATGATTTCTTGAAGTATTAAAGCCCATGAATTAGATTTATTC 626

QY 130 LysHisIleSerSerAsnGluTyrIleIleGluAspSerPheLysLeuLeuAsnSerGlu 149

Db 627 AAAGATTAAAGTACCAACAATATGTTATTAGAATCCATATCAATTATTAGATGATGAT 686

Search completed: May 19, 2003, 16:14:54

Job time : 1292 secs

FEATURES

source

Qy 178 YEKVLAKYKDDLSIKKVIKEEKEKFPSPPTTPPSPAKTDEQKESKFLPFLTNIETL 237
Db 121 YEKVLAKYKDDLSIKKVIKEEKEKFPSPPTTPPSPAKTDEQKESKFLPFLTNIETL 180
Qy 238 YNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIKK 297
Db 181 YNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIKK 240
Qy 298 LINDTKKMDLGLKLLSTGLVQNFNTIISKLEIEGKFQDMLNISQHCYKVKQCPENSGCFR 357
Db 241 LINDTKKMDLGLKLLSTGLVQNFNTIISKLEIEGKFQDMLNISQHCYKVKQCPENSGCFR 300
Qy 358 HLDRECECKLLNYKQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSSRKKITCECTK 417
Db 301 HLDRECECKLLNYKQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSSRKKITCECTK 360
Qy 418 PDSYPLFDGIFCSS 431
Db 361 PDSYPLFDGIFCSS 374

RESULT 2

US-10-098-514-14
; Sequence 14, Application us/10098514
; Publication No. US20020194648A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra P
; APPLICANT: Christopher, David A
; APPLICANT: Vine, Benjamin
; APPLICANT: Su, Wei-Wen
; APPLICANT: Bugos, Robert
; TITLE OF INVENTION: PLASMODIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PRODUCE
; FILE REFERENCE: A-71339/RFT/TAL/NBC
; CURRENT APPLICATION NUMBER: US/10/098,514
; CURRENT FILING DATE: 2002-08-06
; PRIOR APPLICATION NUMBER: US 09/500,376
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: US 60/274,599
; PRIOR FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
US-10-098-514-14

Query Match 84.7%; Score 1928; DB 9; Length 402;

Best Local Similarity 98.1%; Pred. No. 4.1e-115;
Matches 368; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 56 TMAISVTMDNLTSGFENEYDVIYKPLAGVYRSLLKQIEKNIFTNLDNLINSLRKKR 115
Db 20 TAAISVTMDNLTSGFENEYDVIYKPLAGVYRSLLKQIEKNIFTNLDNLINSLRKKR 79
Qy 116 KYFLDVLESQDMQFKHISSENYIIEDSFKLLNSEQKNTLLSKYKIKESVENDIKFAQEG 175
Db 80 KYFLDVLESQDMQFKHISSENYIIEDSFKLLNSEQKNTLLSKYKIKESVENDIKFAQEG 139
Qy 176 ISYEKVLAKYKDDLSIKKVIKEEKEKFPSPPTTPPSPAKTDEQKESKFLPFLTNIE 235
Db 140 ISYEKVLAKYKDDLSIKKVIKEEKEKFPSPPTTPPSPAKTDEQKESKFLPFLTNIE 199
Qy 236 TLYNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIK 295
Db 200 TLYNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFDAI 259
Qy 296 KKLINDTKKMDLGLKLLSTGLVQNFNTIISKLEIEGKFQDMLNISQHCYKVKQCPENSGC 355
Db 260 KKLINDTKKMDLGLKLLSTGLVQNFNTIISKLEIEGKFQDMLNISQHCYKVKQCPENSGC 319

Qy 356 FRHLDRECECKLLNYKQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSSRKKITCEC 415
Db 320 FRHLDRECECKLLNYKQEGDKCVENPNPTCNENNGCGDADAKTCTEEDSGSGNKITCEC 379
Qy 416 TKPDSYPLFDGIFCS 430
Db 380 TKPDSYPLFDGIFCS 394

RESULT 3

US-10-098-514-4
; Sequence 4, Application us/10098514
; Publication No. US20020194648A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra P
; APPLICANT: Christopher, David A
; APPLICANT: Vine, Benjamin
; APPLICANT: Su, Wei-Wen
; APPLICANT: Bugos, Robert
; TITLE OF INVENTION: PLASMODIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PROD
; FILE REFERENCE: A-71339/RFT/TAL/NBC
; CURRENT APPLICATION NUMBER: US/10/098,514
; CURRENT FILING DATE: 2002-08-06
; PRIOR APPLICATION NUMBER: US 09/500,376
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: US 60/274,599
; PRIOR FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
; NAME/KEY: MISC_FEATURE
; LOCATION: (380)..(380)
; OTHER INFORMATION: "Xaa" at position 380 represents a stop codon
US-10-098-514-4

Query Match 84.0%; Score 1911; DB 9; Length 383;

Best Local Similarity 98.1%; Pred. No. 4.7e-114;
Matches 363; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 61 VTMDNLTSGFENEYDVIYKPLAGVYRSLLKQIEKNIFTNLDNLINSLRKKRYFLD 120
Db 6 ITMDNLTSGFENEYDVIYKPLAGVYRSLLKQIEKNIFTNLDNLINSLRKKRYFLD 65
Qy 121 VLESQDMQFKHISSENYIIEDSFKLLNSEQKNTLLSKYKIKESVENDIKFAQEGISYVE 180
Db 66 VLESQDMQFKHISSENYIIEDSFKLLNSEQKNTLLSKYKIKESVENDIKFAQEGISYVE 125
Qy 181 KVLAKYKDDLSIKKVIKEEKEKFPSPPTTPPSPAKTDEQKESKFLPFLTNIETLYNN 240
Db 126 KVLAKYKDDLSIKKVIKEEKEKFPSPPTTPPSPAKTDEQKESKFLPFLTNIETLYNN 185
Qy 241 LVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIKKLN 300
Db 186 LVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHNDFDAIKKLN 245
Qy 301 DDTKKMDLGLKLLSTGLVQNFNTIISKLEIEGKFQDMLNISQHCYKVKQCPENSGCFRHL 360
Db 246 DDTKKMDLGLKLLSTGLVQNFNTIISKLEIEGKFQDMLNISQHCYKVKQCPENSGCFRHL 305
Qy 361 ERECECKLLNYKQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSSRKKITCECTKPD 420
Db 306 ERECECKLLNYKQEGDKCVENPNPTCNENNGCGDADAKTCTEEDSGSGNKITCECTKPD 365
Qy 421 YPLFDGIFCS 430
Db 421 YPLFDGIFCS 430

Db 366 YPLFDGIFCS 375

RESULT 4

US-10-098-514-2

; Sequence 2, Application US/10098514

; Publication No. US20020194648A1

; GENERAL INFORMATION:

; APPLICANT: Chang, Sandra P

; APPLICANT: Christopher, David A

; APPLICANT: Vine, Benjamin

; APPLICANT: Su, Wei-wen

; APPLICANT: Bugeis, Robert

; TITLE OF INVENTION: PLASMODIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PRODUCE

; FILE REFERENCE: A-71339/RET/TAL/NBC

; CURRENT APPLICATION NUMBER: US/10/098,514

; PRIOR FILING DATE: 2002-08-06

; PRIOR APPLICATION NUMBER: US 09/500,376

; PRIOR FILING DATE: 2000-02-08

; PRIOR APPLICATION NUMBER: US 60/274,599

; PRIOR FILING DATE: 2001-03-09

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2

; LENGTH: 383

; TYPE: PRT

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: synthetic

; NAME/KEY: MISC_FEATURE

; LOCATION: (380)..(380)

; OTHER INFORMATION: "Xaa" at position 380 represents a stop codon

US-10-098-514-2

Query Match 83.5%; Score 1900; DB 9; Length 383;

Best Local Similarity 97.3%; Pred. No. 2.3e-113;

Matches 361; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 60 SVTMDNLGSGFENEYDVIYKPLAGYVRSKQIKQIENIETFNLDILNSRLKRRKYFL 119

Db 5 NILSONILSGFENEYDVIYKPLAGYVRSKQIKQIENIETFNLDILNSRLKRRKYFL 64

QY 120 DVLESOLMDFKHSSNEYIIESFLLNSEQKNTLLSKYIKESVENDIKFAQEGISY 179

Db 65 DVLESOLMDFKHSSNEYIIESFLLNSEQKNTLLSKYIKESVENDIKFAQEGISY 124

QY 180 EKVLAQYKDDLESIKKVIKEEKEKFPSPPTTPPSAKTDEQKESKFLPFLNIETLYN 239

Db 125 EKVLAQYKDDLESIKKVIKEEKEKFPSPPTTPPSAKTDEQKESKFLPFLNIETLYN 184

QY 240 NLVNTKIDDLINLKAINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDPEATKKLI 299

Db 185 NLVNTKIDDLINLKAINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNHDFDAKKLI 244

QY 300 NDDTKRDMLGKLLSTGLVGNFPNTIISKIEGFDQMLNISQHCYKQCPENSGCFRHL 359

Db 245 NDDTKRDMLGKLLSTGLVGNFPNTIISKIEGFDQMLNISQHCYKQCPENSGCFRHL 304

QY 360 DERECKCLLNKYQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSRKKITCECTKPD 419

Db 305 DERECKCLLNKYQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSRKKITCECTKPD 364

QY 420 SYPLFDGIFCS 430

Db 365 SYPLFDGIFCS 375

RESULT 5

US-10-087-464-10

; Sequence 10, Application US/10087464

; Publication No. US20030059436A1

; GENERAL INFORMATION:

; APPLICANT: Chishti, Athar

; APPLICANT: Oh, Steven

; APPLICANT: Liu, David

; APPLICANT: Goel, Vikas

; APPLICANT: Li, Xuerong

; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses There

; FILE REFERENCE: SI237/7019

; CURRENT APPLICATION NUMBER: US/10/087,464

; CURRENT FILING DATE: 2002-03-01

; PRIOR APPLICATION NUMBER: US 06/272,930

; PRIOR FILING DATE: 2001-03-02

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 10

; LENGTH: 1639

; TYPE: PRT

; ORGANISM: Plasmodium falciparum

US-10-087-464-10

Query Match 51.6%; Score 1175; DB 9; Length 1639;

Best Local Similarity 56.6%; Pred. No. 1.5e-66;

Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;

QY 34 PDLGTDDDDKAMADIGSIEGRGTMAISVTMDNLSGFENYDVIYKPLAGYVRSKQI 93

Db 1244 PIFGESEED--YDDLGVVVTGEAVTPSV-IDNLSKIENEYEVLYLKLPLAGYVRSKQI 1300

QY 94 EKNIFTNLDILNSRLKRRKYFLDVLSDLMQKHSSNEYIIESFLLNSEQKNT 153

Db 1301 ENNVMTFNVNVDILNSRFNKRNFNVLSDLPYKDTSSNVVVKDPYKFLNKEKRD 1360

QY 154 LKSKYIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKFPSPPTTP 213

Db 1361 FLSSYVNIKSDIDTDFINADVILGYKILSEKYSKLDLSIKKI----- 1404

QY 214 SPAKTDEQKESKFLPFLNIETLYNVLNKKIDDYLNKAKINDCNVEKDEAHVKITKL 273

Db 1405 ----NDKQENKYLPLFNNIETLYNTVNDKIDLFVHLEAKVLNLYTEKSNVEVKIKEL 1460

QY 274 SLDKATDDKIDLFKNPYDPEAIAKKLINDDTFKQMLGKLLSTGLV-QNFPNTIISKIEGK 332

Db 1461 NYLKTQDLKADPKNNFVGIADLSTDYHNHNLTKFLSTGMVFENLAKTVLSNLDGN 1520

QY 333 FQDMLNISQHCYKQCPENSGCFRHLDERECKCLLNKYQEGDKCVENPNPTCNENNGG 392

Db 1521 LQGLMLNISQHCYKQCPQNSGCFRHLDERECKCLLNKYQEGDKCVENPNPTCNENNGG 1580

QY 393 CDADATCTEEDSGSRKKITCECTKPDSPFLFDGIFCS 431

Db 1581 CDADAKCTEEDSGSGNKKITCECTKPDSPFLFDGIFCS 1619

RESULT 6

US-09-978-756-2

; Sequence 2, Application US/09978756

; Patent No. US20020160017A1

; GENERAL INFORMATION:

; APPLICANT: Holder, Anthony

; APPLICANT: Birdsell, Berry

; APPLICANT: Feeney, James

; APPLICANT: Morgan, William

; APPLICANT: Syed, Shabih

; TITLE OF INVENTION: Malaria Vaccine

; FILE REFERENCE: 18396/1005

; CURRENT APPLICATION NUMBER: US/09/978,756

; CURRENT FILING DATE: 2001-10-16

; PRIOR APPLICATION NUMBER: PCT/GB00/01558

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: 09/311,817

; PRIOR FILING DATE: 1999-05-13

; PRIOR APPLICATION NUMBER: 2,271,451

; PRIOR FILING DATE: 1999-05-25

; PRIOR APPLICATION NUMBER: 9909072.2
; PRIOR FILING DATE: 1999-04-20
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-978-756-2

Query Match 51.2%; Score 1165.5; DB 9; Length 376;
Best Local Similarity 59.2%; Pred. No. le-66;
Matches 219; Conservative 54; Mismatches 76; Indels 21; Gaps 2;

Qy 63 MDNLSGFENEYDVLYLKLPLAGVYRSLLKQKQENNNVMTFVNVKDIILSRFKRNFKNVL 122
Db 7 IDNLSKIENEYEVLYLKLPLAGVYRSLLKQKQENNNVMTFVNVKDIILSRFKRNFKNVL 66
Qy 123 ESDLMOFKHISSEYIIEDSFKLNSKQNTLLKSYKIKESVENDIKFAOEGISYYEKV 182
Db 67 ESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSYNIKDSIDTIDINFANDVILGYKIL 126
Qy 183 LAKYKDDLESIKKVIKEEKFPSPPTPPSPAKTDQKKESKFLPFLTNIETLYNNLV 242
Db 127 SEKYKSDLDLSIKYI-----NDKQGENEYLPFLNNIETLYKTVN 166
Qy 243 NKIDYLLINKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFPKNPYDFFAIKKLIND 302
Db 167 DKIDLFIHLEAKVLYTYEKSNEVEVKIKELNYLKTIOQKLADFKKNNFVGIADLSTDY 226
Qy 303 TKKDWLGLLSTGLV-QNFPNTIISKLEGFQDMLNLSOHCVKQKQPCNSGCGFRHLDE 361
Db 227 NHNLLTKFLSTGWFENLAKTVLSNLDGMLNLSOHCVKQKQPCNSGCGFRHLDE 286
Qy 362 REECKCLLYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRRKKITCECTKPDY 421
Db 287 REECKCLLYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDY 346
Qy 422 PLFDGIFCSS 431
Db 347 PLFDGIFCSS 356

RESULT 7

US-10-087-464-34
; Sequence 34, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chishtli, Athar
; APPLICANT: Liu, Steven
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuerong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
; FILE REFERENCE: S1237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-087-464-34

Query Match 51.2%; Score 1165.5; DB 9; Length 376;
Best Local Similarity 59.2%; Pred. No. le-66;
Matches 219; Conservative 54; Mismatches 76; Indels 21; Gaps 2;
Qy 63 MDNLSGFENEYDVLYLKLPLAGVYRSLLKQKQENNNVMTFVNVKDIILSRFKRNFKNVL 122
Db 7 IDNLSKIENEYEVLYLKLPLAGVYRSLLKQKQENNNVMTFVNVKDIILSRFKRNFKNVL 66
Qy 123 ESDLMOFKHISSEYIIEDSFKLNSKQNTLLKSYKIKESVENDIKFAOEGISYYEKV 182
Db 67 ESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSYNIKDSIDTIDINFANDVILGYKIL 126
Qy 183 LAKYKDDLESIKKVIKEEKFPSPPTPPSPAKTDQKKESKFLPFLTNIETLYNNLV 242
Db 127 SEKYKSDLDLSIKYI-----NDKQGENEYLPFLNNIETLYKTVN 166
Qy 243 NKIDYLLINKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFPKNPYDFFAIKKLIND 302
Db 167 DKIDLFIHLEAKVLYTYEKSNEVEVKIKELNYLKTIOQKLADFKKNNFVGIADLSTDY 226
Qy 303 TKKDWLGLLSTGLV-QNFPNTIISKLEGFQDMLNLSOHCVKQKQPCNSGCGFRHLDE 361
Db 227 NHNLLTKFLSTGWFENLAKTVLSNLDGMLNLSOHCVKQKQPCNSGCGFRHLDE 286
Qy 362 REECKCLLYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRRKKITCECTKPDY 421
Db 287 REECKCLLYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDY 346
Qy 422 PLFDGIFCSS 431
Db 347 PLFDGIFCSS 356

RESULT 8

US-10-087-464-11
; Sequence 11, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chishtli, Athar
; APPLICANT: Liu, Steven
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuerong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses There
; FILE REFERENCE: S1237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-087-464-11

Query Match 51.2%; Score 1165.5; DB 9; Length 378;
Best Local Similarity 59.2%; Pred. No. le-66;
Matches 219; Conservative 54; Mismatches 76; Indels 21; Gaps 2;

Qy 63 MDNLSGFENEYDVLYLKLPLAGVYRSLLKQKQENNNVMTFVNVKDIILSRFKRNFKNVL 122
Db 9 IDNLSKIENEYEVLYLKLPLAGVYRSLLKQKQENNNVMTFVNVKDIILSRFKRNFKNVL 68
Qy 123 ESDLMOFKHISSEYIIEDSFKLNSKQNTLLKSYKIKESVENDIKFAOEGISYYEKV 182
Db 67 ESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSYNIKDSIDTIDINFANDVILGYKIL 128
Qy 183 LAKYKDDLESIKKVIKEEKFPSPPTPPSPAKTDQKKESKFLPFLTNIETLYNNLV 242
Db 127 SEKYKSDLDLSIKYI-----NDKQGENEYLPFLNNIETLYKTVN 168
Qy 243 NKIDYLLINKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFPKNPYDFFAIKKLIND 302
Db 169 DKIDLFIHLEAKVLYTYEKSNEVEVKIKELNYLKTIOQKLADFKKNNFVGIADLSTDY 228
Qy 303 TKKDWLGLLSTGLV-QNFPNTIISKLEGFQDMLNLSOHCVKQKQPCNSGCGFRHLDE 361
Db 227 NHNLLTKFLSTGWFENLAKTVLSNLDGMLNLSOHCVKQKQPCNSGCGFRHLDE 286
Qy 362 REECKCLLYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRRKKITCECTKPDY 421
Db 287 REECKCLLYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDY 346
Qy 422 PLFDGIFCSS 431
Db 347 PLFDGIFCSS 356

RESULT 11
US-09-134-333-11
; Sequence 11, Application US/09134333
; Patent No. US20020076403A1
; GENERAL INFORMATION:
; APPLICANT: LONGACRE-ANDRE, SHIRLEY
; APPLICANT: ROTH, CHARLES
; APPLICANT: NATO, FARIDABANO
; APPLICANT: BARNWELL, JOHN
; APPLICANT: MENDIS, KAMINI
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF
; FILE REFERENCE: 0660-0135-0XCIP
; CURRENT APPLICATION NUMBER: US/09/134,333
; EARLIER FILING DATE: 1999-04-18
; EARLIER APPLICATION NUMBER: PCT/FR97/00290
; EARLIER FILING DATE: 1997-02-14
; EARLIER APPLICATION NUMBER: FR96/01822
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Plasmodium cynomolgi.
; FEATURE:
; OTHER INFORMATION: Amino Acids 1-139- REGION I
; FEATURE:
; OTHER INFORMATION: Amino Acids 140-177-REGION II
; FEATURE:
; OTHER INFORMATION: Amino Acids 178-282-REGION III
; FEATURE:
; OTHER INFORMATION: Amino Acids 283-379-REGION IV
US-09-134-333-11

Query Match 32.3%; Score 734.5; DB 10; Length 379;
Best Local Similarity 41.9%; Pred. No. 2.4e-39;
Matches 156; Conservative 74; Mismatches 115; Indels 27; Gaps 10;
Qy 72 NEYDVYIKPLAGVVRSLKQIEKNIFTEFLNLDNLSRLKRRKYFDVLESDLMOFKH 131
Db 22 NEYDVYIKPLAGWKTKIKKPLENHVNALNTIIDLMSRLKRRNYFDVLSNLPYSI 81
Qy 132 ISSNEYIIEFSLNSQKTLKSYKIKESVENDIKFAQEGISYKYLAKYKDDLE 191
Db 82 PHSGEYIKDPYKLLDEKKK-LLGSYKYGASVDKDMVTANDGLAYYQKMGDLYKKHLD 140
Qy 192 SIKKVIKE-----EKEKFPSPPTTSPAKTDQKKE-SKFLPFLNLTETLYNNL 241
Db 141 EVNACIKEVEANINKHDEEIKKIGSEASKANDKNQNAKKELOKYLPLSLSIQKEYSTL 200
Qy 242 VNKIDDIYLNKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIKKLND 301
Db 201 VNKVHSYTDTLKINNCQIEKKTETITVNLKEDYSKDELDVYKQS-----KK--ED 252
Qy 302 DTKKD-MLGKLLSTGLV-QNFPNTIISKIEGKFDQMLNI-SHQCVKKQKCPENSGCFRH 358
Db 253 DVKSSGLEKLMNSKLINQESKALSSELLNVQTO-MLNMSSEHRCIDTNPVNAACVRY 311
Qy 359 LDRECECKLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEDSGSRKKITCECTKP 418
Db 312 LDGTEWRCLLYFXEDAGKCVAPAPNMTCKDKNGGCAPEACKMND----KNEIVCKCTKE 367
Qy 419 DSYPLFDGIFCS 430
Db 368 CSEPLFEGVFCS 379
RESULT 12
US-09-134-333-10
; Sequence 10, Application US/09134333
; Patent No. US20020076403A1

; GENERAL INFORMATION:
; APPLICANT: LONGACRE-ANDRE, SHIRLEY
; APPLICANT: ROTH, CHARLES
; APPLICANT: NATO, FARIDABANO
; APPLICANT: BARNWELL, JOHN
; APPLICANT: MENDIS, KAMINI
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF
; FILE REFERENCE: 0660-0135-0XCIP
; CURRENT APPLICATION NUMBER: US/09/134,333
; EARLIER FILING DATE: 1999-04-18
; EARLIER APPLICATION NUMBER: PCT/FR97/00290
; EARLIER FILING DATE: 1997-02-14
; EARLIER APPLICATION NUMBER: FR96/01822
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-134-333-10
Query Match 23.2%; Score 527; DB 10; Length 108;
Best Local Similarity 94.8%; Pred. No. 8.3e-27;
Matches 91; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Qy 335 DMLNISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCD 394
Db 13 DEFNISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCD 72
Qy 395 ADATCTEDSGSRKKITCECTKPDSPYPLFDGIFCS 430
Db 73 ADAKTEBDSGNGKKITCECTKPDSPYPLFDGIFCS 108

RESULT 13
US-09-134-333-5
; Sequence 5, Application US/09134333
; Patent No. US20020076403A1
; GENERAL INFORMATION:
; APPLICANT: LONGACRE-ANDRE, SHIRLEY
; APPLICANT: ROTH, CHARLES
; APPLICANT: NATO, FARIDABANO
; APPLICANT: BARNWELL, JOHN
; APPLICANT: MENDIS, KAMINI
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF
; FILE REFERENCE: 0660-0135-0XCIP
; CURRENT APPLICATION NUMBER: US/09/134,333
; EARLIER FILING DATE: 1999-04-18
; EARLIER APPLICATION NUMBER: PCT/FR97/00290
; EARLIER FILING DATE: 1997-02-14
; EARLIER APPLICATION NUMBER: FR96/01822
; EARLIER FILING DATE: 1996-02-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 116
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-134-333-5
Query Match 23.2%; Score 527; DB 10; Length 116;
Best Local Similarity 96.8%; Pred. No. 9e-27;
Matches 91; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 338 NISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCDADA 397
Db 3 NISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCDADA 62

QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431
Matches 90; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 63 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSS 96

RESULT 14

US-09-978-756-1

; Sequence 1, Application US/09978756

; Patent No. US20020160017A1

; GENERAL INFORMATION:

; APPLICANT: Holder, Anthony

; APPLICANT: Birdsall, Berry

; APPLICANT: Feeney, James

; APPLICANT: Morgan, William

; APPLICANT: Syed, Shabih

; TITLE OF INVENTION: Malaria Vaccine

; FILE REFERENCE: 18396/1005

; CURRENT APPLICATION NUMBER: US/09/978,756

; PRIOR FILING DATE: 2001-10-16

; PRIOR APPLICATION NUMBER: PCT/GB00/01558

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: 09/311,817

; PRIOR FILING DATE: 1999-05-13

; PRIOR APPLICATION NUMBER: 2,271,451

; PRIOR FILING DATE: 1999-05-25

; PRIOR APPLICATION NUMBER: 9909072.2

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1

; LENGTH: 96

; TYPE: PRT

; ORGANISM: Plasmodium falciparum

US-09-978-756-1

Query Match 23.0%; Score 524; DB 9; Length 96;

Best Local Similarity 95.7%; Pred. No. 1.1e-26;

Matches 90; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 338 NISOHCQVKKQCPNSGCFRHLDERECKLLNKKQBGDKCVENPNTCTENNGGCCDADA 397

Db 1 NISOHCQVKKQCPNSGCFRHLDERECKLLNKKQBGDKCVENPNTCTENNGGCCDADA 60

QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431

Matches 90; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 61 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSS 94

RESULT 15

US-10-087-464-35

; Sequence 35, Application US/10087464

; Publication No. US20030059436A1

; GENERAL INFORMATION:

; APPLICANT: Chishti, Athar

; APPLICANT: Oh, Steven,

; APPLICANT: Liu, David,

; APPLICANT: Goel, Vikas

; APPLICANT: Li, Xuerong

; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof

; FILE REFERENCE: S1237/7019

; CURRENT APPLICATION NUMBER: US/10/087,464

; PRIOR FILING DATE: 2002-03-01

; PRIOR APPLICATION NUMBER: US 06/272,930

; PRIOR FILING DATE: 2001-03-02

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 35

; LENGTH: 114

; TYPE: PRT

; ORGANISM: Plasmodium falciparum

US-10-087-464-35

Query Match 23.0%; Score 524; DB 9; Length 114;

Best Local Similarity 95.7%; Pred. No. 1.4e-26;
Matches 90; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 338 NISOHCQVKKQCPNSGCFRHLDERECKLLNKKQBGDKCVENPNTCTENNGGCCDADA 397

Db 1 NISOHCQVKKQCPNSGCFRHLDERECKLLNKKQBGDKCVENPNTCTENNGGCCDADA 60

QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431

Matches 90; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 61 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSS 94

RESULT 16

US-09-134-333-2

; Sequence 2, Application US/09134333

; Patent No. US20020076403A1

; GENERAL INFORMATION:

; APPLICANT: LONGACRE-ANDRE, SHIRLEY

; APPLICANT: ROTH, CHARLES

; APPLICANT: NATO, FARIDABANO

; APPLICANT: BARNWELL, JOHN

; APPLICANT: MENDIS, KAMINI

; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF

; FILE REFERENCE: 0660-0135-0XCIP

; CURRENT APPLICATION NUMBER: US/09/134,333

; PRIOR FILING DATE: 1999-04-18

; EARLIER APPLICATION NUMBER: PCT/FR97/00290

; EARLIER FILING DATE: 1997-02-14

; EARLIER APPLICATION NUMBER: FR96/01822

; EARLIER FILING DATE: 1996-02-14

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 95

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

US-09-134-333-2

Query Match 23.0%; Score 523; DB 10; Length 95;

Best Local Similarity 96.8%; Pred. No. 1.3e-26;

Matches 90; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 338 NISOHCQVKKQCPNSGCFRHLDERECKLLNKKQBGDKCVENPNTCTENNGGCCDADA 397

Db 3 NISOHCQVKKQCPNSGCFRHLDERECKLLNKKQBGDKCVENPNTCTENNGGCCDADA 62

QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 430

Matches 90; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 63 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSS 95

RESULT 17

US-09-134-333-8

; Sequence 8, Application US/09134333

; Patent No. US20020076403A1

; GENERAL INFORMATION:

; APPLICANT: LONGACRE-ANDRE, SHIRLEY

; APPLICANT: ROTH, CHARLES

; APPLICANT: NATO, FARIDABANO

; APPLICANT: BARNWELL, JOHN

; APPLICANT: MENDIS, KAMINI

; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF

; FILE REFERENCE: 0660-0135-0XCIP

; CURRENT APPLICATION NUMBER: US/09/134,333

; PRIOR FILING DATE: 1999-04-18

; EARLIER APPLICATION NUMBER: PCT/FR97/00290

; EARLIER FILING DATE: 1997-02-14

; EARLIER APPLICATION NUMBER: FR96/01822

; EARLIER FILING DATE: 1996-02-14

```

; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-134-333-8

Query Match      23.0%; Score 523; DB 10; Length 127;
Best Local Similarity 96.8%; Pred. No. 1.8e-26;
Matches 90; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 338 NISQHCYKQCPENSGCFRHLDERECKCLLNKQEGKCVENPPTCENNNGCCDADA 397
Db 35 NISQHCYKQCPENSGCFRHLDERECKCLLNKQEGKCVENPPTCENNNGCCDADA 94

Qy 398 TCTEDSGSRKKITCECTKPDSPFLFDGIFCS 430
Db 95 KCTEDSGSNGKKITCECTKPDSPFLFDGIFCS 127

RESULT 18
US-09-134-333-14
; Sequence 14, Application US/09134333
; Patent No. US20020076403A1
; GENERAL INFORMATION:
; APPLICANT: LONGACRE-ANDRE, SHIRLEY
; APPLICANT: ROTH, CHARLES
; APPLICANT: NATO, FARIDABANO
; APPLICANT: BARNWELL, JOHN
; APPLICANT: MENDIS, KAMINI
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF
; FILE REFERENCE: 0660-0135-0XCIP
; CURRENT APPLICATION NUMBER: US/09/134,333
; CURRENT FILING DATE: 1999-04-18
; EARLIER APPLICATION NUMBER: PCT/FR97/00290
; EARLIER FILING DATE: 1997-02-14
; EARLIER APPLICATION NUMBER: FR96/01822
; EARLIER FILING DATE: 1996-02-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:ALIGNMENT
; FEATURE:
; OTHER INFORMATION: Amino Acids 1-115-REGION I
; FEATURE:
; OTHER INFORMATION: Amino Acids 116-125-REGION II
; FEATURE:
; OTHER INFORMATION: Amino Acids 126-197-REGION III
; FEATURE:
; OTHER INFORMATION: Amino Acids 198-281-REGION IV
US-09-134-333-14

Query Match      18.7%; Score 424.5; DB 10; Length 281;
Best Local Similarity 35.7%; Pred. No. 8.3e-20;
Matches 131; Conservative 55; Mismatches 56; Indels 115; Gaps 25;

Qy 74 YDVIYKPLAGYVSLKKQIEKNFTNLNLN-DILNSRLKRRKYFLDVLSDLMQPKHI 132
Db 20 YDVIY-KPLAGMYTKIK-LENHV---NANTNIDMLDSALKKANYFL-VLNSDLN----- 68

Qy 133 SSNEYIIDSFKLLNSQKNTLLSKYIKESVENDIKFAQEGISYVEKVLAKYKDDLES 192
Db 69 PSGEYIINKDYPKLLDLEKKK-LGSYKYGASDDT---ANDG--YYKMGLYKH----- 114

Qy 193 IKKVIKEKEKFPSPPTTPSPAKTDOKKESKFLPLFTNIETL---YNNLVNKIDIDYL 249
Db 115 ---LVKVEID-----KKGKAKELKYLFLSQKEYLKVTDL----- 149
```

```

Qy 250 INLKAKINDCNVEKDEAHVKITKLSDLKAIDDKID-LFKNPYDFEAIKKLINDDTKKDML 308
Db 150 ---KINNCKEKEEV-----KLDYKNDELYKS----- 174

Qy 309 GKLLSTGLVQNPNTIISKLEGFQDMLNI-----SOHCYKQCPENSGCFRHLDERE 363
Db 175 -KVSSGLLEK-----LMSKLIESKL-SLLNVQTQLMSSEHCITDTNVP-NAACYRYLDGTE 227

Qy 364 ECKCLLNKQEGKCVENPPTCENNNGCCDADATCTEEDSGSRKKITCECTKPDSPYL 423
Db 228 EWRCLL-FKE--GKVC--PANTC-KDNGGCAPEAECKMDN-----IVCKCTREGSEPL 274

Qy 424 FDGIFCS 430
Db 275 FEGVFCs 281

RESULT 19
US-09-804-626-4
; Sequence 4, Application US/09804626
; Patent No. US20020128190A1
; GENERAL INFORMATION:
; APPLICANT: Lobel, Leslie
; APPLICANT: Lustbader, Joyce
; TITLE OF INVENTION: EXPRESSION OF PROPERLY FOLDED AND SOLUBLE EXTRACELLULAR DOMAIN
; FILE REFERENCE: 0575/62259/JPW/SHS
; CURRENT APPLICATION NUMBER: US/09/804,626
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 516
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-804-626-4

Query Match      13.5%; Score 306.5; DB 10; Length 516;
Best Local Similarity 27.2%; Pred. No. 5.3e-12;
Matches 124; Conservative 63; Mismatches 156; Indels 113; Gaps 19;

Qy 1 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDDDDKAMADIGSIEGRGTMAIS 60
Db 116 MHHHHHSSGLVPRGSGMKETAATAAKFERQHMDSPDLGTDDDDKAMADIGS-EFRVFLCOE 174

Qy 61 VTMDNILSGFENEYDVIYKPLAGYVSLKK-----QIEKNFTFNLNLDILNSRLK 113
Db 175 SKVTEIPSDLPR--NATELRFVLTKLVIQKAFSGFDLEK---IEISQNDV----- 222

Qy 114 KRKYFLDVLESDL-----MQFKHSSNEYIIDSFKLLNSEQ-----KNTLLKSYKY 160
Db 223 -----LEVIEADVSNLPKLEIRKANNLLYITPEAFQNLPLNQLYLLISNTGKHLPD 277

Qy 161 IKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAKTDE 220
Db 278 VHK-----IHSLOKVLIDQDNI---NIHTIERNSEFVG----- 307

Qy 221 QKESKFLPLFTN-IETLYNNLVNKIDIDYILNKLAKINDCNVEKDEAH-----VKI 270
Db 308 LSFESVILWLNKNGIOETHNCAFNGTQLDVNLSDNNNLEELPNDVHFAGSGPVILDISR 367

Qy 271 TKLSDLKAIDDKIDLKFNPDYFAIKKLINDDTKKDMLGKLLSTGLVQNPNTIISKLEIE 330
Db 368 TRIHSLPS-----YGLENLKLL-----RARSTVNLKKLPT--LEKLV- 402

Qy 331 GKFOFMLNISQHCY-----KQCPENSGCFRHLDERECKCLLNKQEGKCVENPNTC 386
Db 403 ALMEASLTYPSSHCAFAFNWRQISLHPICNKSILRQEVDMYMTQAGQORSSLAED----- 457

Qy 387 NENNGCCDADATCTEEDSGSRKKITCECT-KPDYS 421
Db 458 NESSYSRGFDWYTFEYDIDLNCNEVVVDVTCSPKPAF 493
```



```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1011
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-1011

Query Match      11.5%; Score 262.5; DB 9; Length 355;
Best Local Similarity 49.2%; Pred. No. 2.1e-09;
Matches 60; Conservative 10; Mismatches 19; Indels 33; Gaps 5;

QY 1 MHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAMADIGSGRGTMAL- 59
   |||||
Db 116 MHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAMDRL--VOREGTRAVY 173
   |||||
QY 60 -----SVTMDNLGTFENEYDVIYLP--LAGVYRSLKKQTEKN 96
   |||||
Db 174 LASVAAPFPAAGATCLSHSVAVVTASAALTGF--TFSAQLILPYTLASLYHR-----EKQ 226
   |||||
QY 97 IF 98
   :|
Db 227 VF 228

RESULT 23
US-09-681-938-1
; Sequence 1, Application US/09681938
; Publication No. US20030003584A1
; GENERAL INFORMATION:
; APPLICANT: Adams, Darius
; TITLE OF INVENTION: Liposomal Vector Binding Protein for Hepatocyte DNA Delivery
; FILE REFERENCE: 29147
; CURRENT APPLICATION NUMBER: US/09/681,938
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: 09/681,938
; PRIOR FILING DATE: 2001-06-28
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Custom Liposomal Vector Binding Protein for Hepatocyte DNA Delivery
; FEATURE:
; NAME/KEY: BINDING
; LOCATION: (1)..(201)
US-09-681-938-1

Query Match      11.4%; Score 258.5; DB 9; Length 201;
Best Local Similarity 31.6%; Pred. No. 1.9e-09;
Matches 83; Conservative 27; Mismatches 70; Indels 83; Gaps 9;

QY 1 MHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAMADIGSGRGTMALIS 60
   |||||
Db 1 MHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAMA----- 46

QY 61 VTMDNLGTFENEYDVIYLPKPLAGVYRSLKKQIEKNFTFNLDNLNLSLKKRKYFLD 120
   |||||
Db 47 -----SPKRSFKRSKRSKRGIPY---LSE-----D 71

QY 121 VLESDLMQFKHISNEYIIDS--FKLLNSQKN-----TLKLSYKIKESVENDIK 170
   |||||
Db 72 ELKAAEAFKRNHTEFVGHDSGSESTVSGQSPDPQVQGITWTKDLKANRGKM--DVS 129

QY 171 FAQEGI-----SYEYKVLAKYKDDLESIKKVIKEKEKFPSPPTTPPSPAK-TDEQKES 225
   |||||
Db 130 GVQAPVGAITTIEDPVLAK-----KVPETFPELKPGEISRHTSDHMSIY 172

QY 226 KFLPLFTNIETLNNLYNKIDDY 248
   |||||
Db 173 KFMGRSHFLCTFTFNSNNKEY 195

RESULT 24
```

```
US-10-042-945-28
; Sequence 28, Application US/10042945
; Publication No. US20030045468A1
; GENERAL INFORMATION:
; APPLICANT: Fling, Steven P.
; APPLICANT: Foy, Teresa M.
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Wang, Aijun
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Sutherland, R. Alec.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY,
; TITLE OF INVENTION: DIAGNOSIS AND MONITORING OF BREAST CANCER
; FILE REFERENCE: 210121.479C3
; CURRENT APPLICATION NUMBER: US/10/042,945
; CURRENT FILING DATE: 2002-01-08
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-042-945-28

Query Match      7.3%; Score 167; DB 9; Length 55;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GSGMKETAAAKFERQHMDSPDLGTDDDDKAMA 46
   |||||
Db 1 GSGMKETAAAKFERQHMDSPDLGTDDDDKAMA 32

RESULT 25
US-09-757-417-28
; Sequence 28, Application US/09757417
; Patent No. US20020082216A1
; GENERAL INFORMATION:
; APPLICANT: Fanger, Gary R.
; APPLICANT: Foy, Theresa M.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY, DIAGNOSIS AND MONITORING OF BREAST CANCER
; FILE REFERENCE: 210121.479C1
; CURRENT APPLICATION NUMBER: US/09/757,417
; CURRENT FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-757-417-28

Query Match      7.3%; Score 167; DB 10; Length 55;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GSGMKETAAAKFERQHMDSPDLGTDDDDKAMA 46
   |||||
Db 1 GSGMKETAAAKFERQHMDSPDLGTDDDDKAMA 32

RESULT 26
US-09-842-930A-25
; Sequence 25, Application US/09842930A
; Publication No. US20020197681A1
; GENERAL INFORMATION:
; APPLICANT: Weigel, Paul
; TITLE OF INVENTION: Identification of Hyaluronan Receptor for Endocytosis
; FILE REFERENCE: 5820.603
; CURRENT APPLICATION NUMBER: US/09/842,930A
```

;; CURRENT FILING DATE: 2001-04-22
;; PRIOR APPLICATION NUMBER: 60/245,320
;; PRIOR FILING DATE: 2000-11-02
;; PRIOR APPLICATION NUMBER: 60/199,538
;; PRIOR FILING DATE: 2000-04-25
;; NUMBER OF SEQ ID NOS: 56
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 25
;; LENGTH: 1394
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-842-930A-25

Query Match 6.4%; Score 145.5; DB 9; Length 1394;
Best Local Similarity 19.7%; Pred. No. 0.29;
Matches 88; Conservative 77; Mismatches 159; Indels 123; Gaps 22;

QY 50 STEGRCTMAISVTMDNILSGFENEVDYVILKPLAGVYRSLLKQIEKNITFFNLN-----104
DB 8 AIEADAYTVFAPNNNAIENYIREKKVL-----SLEDVLYRHVLEEKLL 53
QY 105 -NDILNSRLKKR-----KYFLD-VLESDDL-----QFKHISSEYIIIESFKLLNSEOK 151
DB 54 KNDLHGMRHRETMLGFSYFLSFELHNDQLYVNEAPINYTNVATDKGVIQ--YNLANAIEA 111
QY 152 NTLKSYKYIKESVENDIKFAQEGISYKVLAKYKDDLESIKKYIKEKEKXPPSPPTT 211
DB 112 ADAYTVFAPNNNAIENYIR-ERKVLSEEDVL-RYHVLE--EKLKNDLH-----158
QY 212 PPSPAKTDOEKESKFLPFLTNIETLYNNLVNKIDIDYLNKAKINDCNVDEAHVKIT 271
DB 159 --NGMHRMETLGFYSFLSFELHNDQLYV-----EAPINYTNVATDKGVIQYN 204
QY 272 KLSDLKAID-----DKI-----DLFKNPYDFAIKKLINDDT-----303
DB 205 LANATEAADAYTVFAPNNNAIENYIREKKVLSLEEDVLR--THVLEEKLLNDLHGWH 262
QY 304 KKDMLG-----KLLSTGLVQNPFTIISKLI-----EGKFDMLNISQHCQVK 346
DB 263 RETMLGFSYFLSFELHNDQLYVNEAPINYTNVATDKGVCAAGFQGGTICTAINACEISN 322
QY 347 KCPENSGCFRHLDBRECKLLNYKQEGDKVE-NPNTCTENNGGCCDADATCEEDSG 405
DB 323 GCSAKADCKRTTPGRVCTKAGYTGOGIVCLEINP-----CLENHGGCDKNAECTQ-----375
QY 406 SSRKKITCECTKPDSPYFLF--DGIFCS 430
DB 376 TGPNOACNC-----LPATYTGDKVCT 397

RESULT 27
US-09-815-242-5251
; Sequence 5251, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727

;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows version 4.0
;; SEQ ID NO 5251
;; LENGTH: 996
;; TYPE: PRT
;; ORGANISM: Staphylococcus aureus
US-09-815-242-5251

Query Match 6.3%; Score 144; DB 10; Length 996;
Best Local Similarity 20.4%; Pred. No. 0.24;
Matches 93; Conservative 70; Mismatches 133; Indels 161; Gaps 21;

QY 25 KPEROHMDSPDLGTDGDD-----DKAMADIGSIEGRGTMAISVTMDNILSGFE 71
DB 387 KQOOSTLNIERLTERTEDEQGLDELNNLKGLEKVEDSIGNESDYKILIE--LNNAITNIN 444
QY 72 NEYDVI-----YLKPLAGVYRSLLKQI--EKNIF-----TFNLNLD-----106
DB 445 NEINVIKENEKAKDELDKLGSQOELENOINEEKTILKNLEIKLDYDKSKLDLNDKESF 504
QY 107 -----LNSRLKKRYFLDY--LESOLMQFKHISSEYIIIESFKLLNSE 149
DB 505 ISEIKSAVIGDQPCGNEIQDLGHIDFDSIAKQNETKEATEANTHTMESIAVHNS 564
QY 150 QKNTLLSKYIKESVEN-DIKFAQEGISYKVLAK-----YKDDLESIKKVIK 198
DB 565 -----IKFVNEKISINIK--TQSDLSL--EVLNKRLEENALNNQDLNKFIEOMK 613
QY 199 EEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL--YNNLV-----242
DB 614 EEKDNLTQLHINKQLRLNKNESELKICRNL--ITFEFETLSKYNNTITFEVDYKYYIQDYN 671
QY 243 -----NKIDDLVILNKAK-----INDCNVEKDEAHVKITKL 273
DB 672 QHQSNSQIETDKLIQISQRKLIEBQNNLHYNQLETYNNDLNEQSIEMEMSRNLITDD 731
QY 274 SDLKAI-----DDKIDLFKNPY-DPE-----ATKKLINDDTKDD-----ML 308
DB 732 NDINEIIAWRGEOEELEQRDRITYKRYHEFEMEIALESITKDKELDDSKKDDYELKK 791
QY 309 GKLLSTGLVQNPFTIISKLEKGFODMLNISQHCQV 345
DB 792 GKM-----NTLIDEVSAVHYOCNNIKKTQSI 818

RESULT 28
US-09-815-242-12141
; Sequence 12141, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12141
; LENGTH: 1009
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12141

Query Match 6.3%; Score 144; DB 10; Length 1009;
Best Local Similarity 20.4%; Pred. No. 0.25;
Matches 93; Conservative 70; Mismatches 133; Indels 161; Gaps 21;

Qy 25 KFERHMDSPDLGTDD-----DKAMADIGSIEGRGTMAISVTMDNLISGFE 71
Db 387 KYQSYLNIERTDEQGLDELNLKLGKVEDSIGNESDYEKIIE--LNNAITNIN 444
Qy 72 NEYDVI-----YLKPLAGVYRSLLKQI--EKNIF-----TFNLNLND---- 106
Db 445 NEINVIKENAKDELKLLGSKOLENQINEEKTILANLEIKLDYDKSKLDLNDKESF 504
Qy 107 -----ILSRLLKRRKYFDV--LESDLMOFKHSSNEYIIEISFKLLNSE 149
Db 505 ISEKSAVKIGDQPCIGCNEIQDLGHDFDSIAKRONKEIKEIANIHTMESNIAVHNS 564
Qy 150 QKNTLLKVKYKESVEN-DIKFAQEGISYKVLAK-----YKDDLESIKKVIK 198
Db 565 -----IKFVNEKISINIK--TQSDLSL--EVLNKRLLNENALNQRLNKFIEQMK 613
Qy 199 BEKEKFPSPPTTPSPAKTDEQKESKFLPLTNIETL--YNNLV----- 242
Db 614 BEKNLTQIHNKQLRLKNESELKICRNL--ITEFETLSKYNNITNFEVDYKVIQDVN 671
Qy 243 -----NKIDDYLLNLKAK-----INDCNVEKDEAHVKITKL 273
Db 672 QHQSNSQIEDKLIQLSQKRIEQNNLHNENQLETYNNDLENEQSIEMEMSRNLTTDD 731
Qy 274 SLDKAI-----DDKIDLFKNPY--DFE-----AIKKLINDDTKDD----ML 308
Db 732 NDINEIINAWRGEQELEQKRDYTKRHYHEFEMETARLESITKDELSDKLKDDYELKK 791
Qy 309 GKLLSTGLVQFPNTIISKLEGFQDMNLNISOHQCV 345
Db 792 GKM-----NTLIDEYSVAVHYQCCNNLKKTKQSI 818

RESULT 29
US-09-888-615-65
; Sequence 65, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26

; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
; LENGTH: 980
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-888-615-65

Query Match 6.2%; Score 140.5; DB 10; Length 980;
Best Local Similarity 20.9%; Pred. No. 0.4;
Matches 77; Conservative 64; Mismatches 154; Indels 73; Gaps 15;

Qy 67 LSGFENEYDVIYKPLAGVYRSLLKQIIEKNIFTNLLNLDILNSRLKRRKYFDLVLESDL 126
Db 341 LOGFSNLGNTCYM-----NAILQSLFSLQSPANDLLKQIPWKKIPLNAL---I 386
Qy 127 MOFKHISSNEYIIEISFKLLSEKQNTLLSKYKIKESVENDIKFAQEGISYI----- 179
Db 387 RRFHALLVKK-----DICNSETKKDLK-----KVKNASATAERFSYGMQDAHE 432
Qy 180 --EKVLAKYKDDLESIKKVIKEKEKFFSPPTTPPSAKTDEQKESKFLPLTNIET 236
Db 433 FLSCQDLQKEDMEKLNKTWKTEPVSGEENSPDISATRAYT-----CPVIITNLEFE 483
Qy 237 LYNLVNKKIDYILNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFE-AI 295
Db 484 VOHSIICKACGEIIPKREQFNLDLIDLRKKPLPP-----RSIQDSLDFPRAEELEYSC 539
Qy 296 KK-----LINDDTYKMDLGLLSTGLVQNPNTIISKLEGKFDMLNLSIQHQCVKKQC 349
Db 540 EKCGGKCALVRH--KFNRLPRVILHLKARYSFNVALS--LNNKIGQCVIIPRYLTSSHC 595
Qy 350 PENS-----GCFRHLDERECKLLNLYKQEGDKCVENPN--PTCENNENGGCDATCTEE 402
Db 596 TENKPPFTLWSAHMAMSRPLKA-----SQMVNSCITSPSTPSKFTFKSKSSLLCLDS 651
Qy 403 DSGSSRKK 410
Db 652 DSEDELKR 659

RESULT 30
US-10-267-311-21
; Sequence 21, Application US/10267311
; Publication No. US20030050469A1
; GENERAL INFORMATION:
; APPLICANT: Siegel, Marvin
; APPLICANT: Chu, N. Randall
; APPLICANT: Mizzen, Lee A.
; TITLE OF INVENTION: INDUCTION OF A THI-LIKE RESPONSE IN VITRO
; FILE REFERENCE: 12071/002001
; CURRENT APPLICATION NUMBER: US/10/267,311
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US/09/613,303
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/143,757
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 948
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion sequence
US-10-267-311-21

Query Match 5.8%; Score 131.5; DB 9; Length 948;
Best Local Similarity 20.0%; Pred. No. 1.4;
Matches 82; Conservative 62; Mismatches 123; Indels 143; Gaps 16;


```
; Sequence 2239, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secretist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C2
; CURRENT APPLICATION NUMBER: US/10/146.502
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2239
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-146-502-2239

Query Match      5.6%; Score 128; DB 9; Length 391;
Best Local Similarity 18.8%; Pred. No. 0.83; Mismatches 171; Indels 120; Gaps 16;
Matches 82; Conservative 64;

Qy 2 HHHHHHSGLVPRGSMKETAATAKFERQHMDSPDLGTDGDDKAMADIGSIEGRGTAISV 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 HHHHHHSGLVPRGSHM-----DARRVPQKDLRVKLNKFKRYVKLI-----SMETSS 52

Qy 62 TMDNLTSGFENEYDVIYKPLAGVYRSLLKQIEKNIFTNLDNLNLSRLKRRKRYFDV 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 53 SSDSCGSFASD-----NFANRLQ-----SV 74

Qy 122 LESDLMQFKHSSNEYIEDSFKLLNSP-QKNTLLKSYKIKESV--ENDIKFAQEGISY 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 75 REGCTRQCRHSGLRVAMKFPARGATNKKRAESQPSNSVTSNDSSEDSGMMF 134

Qy 179 YEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTPSPAKTDEQK----- 223
   || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 135 LEKRALNIKONKAKLKM-SELESFPGSFRGRLPLGSDSQSRPRRTTTPGVASRRNP 193

Qy 224 ESKFLP-----FLTNIETLYNNLVNKIDYLLNLKAKINDCNVEKDEAHVKITKLSDL 276
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 194 ERRARLTRSRRLGSLDALPMEEEDYMLVRKRKTVDGYMNEDD-----LPRS 246

Qy 277 KAIDDKIDLFKNPYDFAIKKLLINDYTKKMLGLLSTGLVQNPNTIISKLEGKFQDM 336
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 247 RRSRSVTL---PHIRPVEITEELE-----NVCSNSREKIYNS 285

Qy 337 LNISSOHCVKQCPENSGCFRHLDEREC-----KCLLN-YKQGGDKCVENPNPTC 386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 286 LGSTCHOCROKTIQKNTNC-----RNPDCMGVGFQCGPCLNRYGVEVRDALLDPNWHC 340

Qy 387 NENNGGCDADATCEED 403
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 341 PPCRGICNC-SFCRORD 356

RESULT 34
US-09-764-864-1160
; Sequence 1160, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764.864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1160
; LENGTH: 580

; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-764-864-1160

Query Match      5.6%; Score 128; DB 10; Length 580;
Best Local Similarity 19.5%; Pred. No. 1.3;
Matches 63; Conservative 64; Mismatches 108; Indels 88; Gaps 15;

Qy 74 YDVIYKPLAGVYRSLLKQIEKNIFTNLDNLNLSRLKRRKRYFDLVLESDL--MOKFK 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 YSMLFLK-----LESQVQAEDEVYLLNLN-----HPKRTILLIKSSWPMDLKI 47

Qy 132 ISSNEYIIEDSFKLLNSEQKNTLLKSYKI-KESVENDIKFAQEGISYIEKVLAKYKDDL 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 48 IEHN-----IRNDKYAGEGMIEDMKLMFRNARHYNEEGSOVYNDA 88

Qy 191 ESIIKKVIEKEKEK---PPSSPTTPPS---PAKTDQEKESKFL-PFLTNIETLYNNLVN 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 89 HILEKLLKRRKELGPLPDDDDMASPKLJSRKSGISPKKSKYKMTPMQOKLNEYEAVKN 148

Qy 244 KID-----DYLNILKAKINDCNVEKDEAHVKITKLSDLKA-IDDK 282
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 149 YTDKGRRLSAIFLRLPSRSELDPYLTIKKPM--DMEKIRSHMMANKYQDIDSWVEDF 205

Qy 283 IDLPKNPYDFAIKKLLIND-----TKDMLGKLLSTGLVQNPNT--IISKLE 330
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 206 VMFNENACTYNEPESLIYKDALVLKLVLETRDLEGEDEDS-----HVPNVTLTIQELIH 260

Qy 331 GKFDMLNISOHCVKQCPENS 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 NLF---VSMVSHQDDEGRCYSDS 280

RESULT 35
US-09-298-523B-61
; Sequence 61, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298.523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-09-298-523B-61

Query Match      5.6%; Score 127.5; DB 9; Length 690;
Best Local Similarity 23.8%; Pred. No. 1.8;
Matches 81; Conservative 47; Mismatches 145; Indels 67; Gaps 13;

Qy 18 MKETAAAKFERQHMDSP-----DLGTDGDDKAMADIGSIEGRGTAISVTDN 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 245 LKEAVKENVATSEQDKPKRAKRGVSGELATPDKKENDAKSSDSSVGEETLPSPLNAN 304

Qy 66 ILSGFENEYDV-IYKPLAGVYRSLLKQIEKNIFTNLDNLNLSRLKRRKRYFDLVLES 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 305 -ESQTEHRKDVDEYIKKMLS-----EIQLDRRKHTQNVNLIKLSAIRTLYELSVLKE 358

Qy 125 D-----LMQFKHISSNEVIEDSFKLL-----NSQKNTLL 155
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 359 NSKKEEUTSKTKAELTAFAEQFKKDTLKPEKKVAEAKKVEEAKKAKDKQEDRRNPT 418

Qy 156 KSYKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVIKEEKEKFPSS-SPPTT 214
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 419 NTKTLEAESDVK-----VKEAELELVKEEANSNEEKIKQAKEKVESKKAETRL 473

Qy 215 PAKTDQEKESKFLPFLTNIETLYNNLVNKIDYLLNLKAKINDCNVEKDEAHVKITKL- 273
```



```

QY 18 MKETAAKPERQHMDSP-----DLGTD--DDKAMADIGSTEGRGTMAISVTMDN 65
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 208 LKEAVEKNVATSQDKPKRKRKRGVSGELATPDKKENDAKSDSSVGETLSPSPUNMAN 267
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 66 ILSGFENEYDV- IYLKPLAGVVRSLKKQJEKNIETFTNLNLDNLILSLRLLKKRYFLDLVLES 124
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

```

Qy	18	MKETAAKAFROHMDSP-----DLGTDD--DDKAMADIGSTEGRGTAISYTMDN	65
Db	245	LKEAVEKNVATSEODPKRRARGVSGELATDPKDENDAKSSDSSVGEETLPSPSLNMAN	304
Qy	66	ILSGFENEYDV-TYLPKLAGVYSLKKQTEKNFTFNLNLDILNRLKRRKYFDVLDES	124
Db	305	-ESQTEHRKDVEYIKMMLS-----ETQLDRRRKHTONVNLIKLSAIKTYLYELSVLKE	358
Qy	125	D-----LMOPKHISSNEYIIEDSFKLL---NSEQKNTLL	155
Db	359	NSKKEELTSKTKAELTAAPFQFKDITLPKPKVAAEAKKVEAKKAKDOKGEDRRNYT	418
Qy	156	KSVKYIK-ESVENDIKFAOEGISYVEKVLAKYKDDLESIKVKIKEKEKPPS-\$PpTTPp	213
Db	419	NTYKTLLETAESOVK-----VKEAEULVKEPANSRNEEKIKQAKEVKEKKABATRL	473
Qy	214	SPAKTDQKKESKFLPLFTNIETLYNNLVNKIDDYLLINKAKINDCNVEKAHVKITKL	273
Db	474	EKIKTDKKAAEBAKRAESEKKAFAKQVDAEYALEAKTAELEYE-----VQRL	526
Qy	274	-SOLKAID--DKIDLFPKNPYDFEAIKKLLINDPTKDMGLKGL	311
Db	527	EKELKETDESDEYTLKEGURAPLOSKL---DTPKAKLSKL	564

Search completed: May 12, 2003, 10:25:59
Job time : 58 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 10:15:38 ; Search time 46 seconds
(without alignments)
900.738 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 2275

Sequence: 1 MHHHHHHSSGLVPRGSMKE.....TCECTKPDSPYPLFDGIFCSS 431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1946.5	85.6	1701	2 A54498	major merozoite su
2	1935.5	85.1	1726	1 SA2QGM	major merozoite su
3	1925.5	84.6	1701	2 A36868	major merozoite su
4	1917	84.3	651	2 S47282	merozoite surface
5	1916.5	84.2	1726	2 A45948	major merozoite su
6	1175	51.6	400	2 A45545	major merozoite su
7	1175	51.6	1639	2 S05603	major merozoite su
8	1051.5	46.2	1631	1 SA2QK1	major merozoite su
9	1007.5	44.3	1640	2 A24594	probable major sur
10	816	35.9	1726	2 A39401	merozoite surface
11	793.5	34.9	1751	2 A45604	major blood-stage
12	556.5	24.5	680	2 A28121	major merozoite su
13	556.5	24.5	1772	2 A45532	major merozoite su
14	552	24.3	1785	2 A45546	major merozoite su
15	180.5	7.9	980	2 E71605	hypothetical prote
16	167.5	7.4	2269	2 T28677	rhoptry protein -
17	166	7.3	2166	2 G70163	hypothetical prote
18	160.5	7.1	1169	2 A64505	pl15 homolog - Met
19	156.5	6.9	1939	2 T18372	repeat organellar
20	154.5	6.8	1156	2 B70356	chromosome assembl
21	152	6.7	1191	2 B97116	chromosome segrega
22	152	6.7	2401	2 T28676	rhoptry protein -
23	149.5	6.6	652	2 B59102	hypothetical prote
24	147	6.5	909	2 C97325	hypothetical prote
25	146.5	6.4	622	2 S51972	SPC72 protein - ye
26	146	6.4	1619	2 T18499	hypothetical prote
27	146	6.4	3724	2 T18427	hypothetical prote
28	145	6.4	1130	2 T34081	hypothetical prote
29	144	6.3	671	2 H64502	hypothetical prote

30	144	6.3	3394	2 T18501	hypothetical prote
31	143.5	6.3	1086	2 S16752	major merozoite su
32	143	6.3	540	2 S38085	hypothetical prote
33	142.5	6.3	864	2 B90395	purine Nriase [imp
34	142.5	6.3	963	2 C90535	conserved hypotet
35	142	6.2	1005	2 A64465	hypothetical prote
36	142	6.2	1187	2 T18355	hypothetical prote
37	142	6.2	1679	2 S48385	hypothetical prote
38	142	6.2	2819	2 A90551	conserved hypotet
39	140.5	6.2	819	2 E70105	pl15 protein homol
40	140	6.2	1009	2 C89910	hypothetical prote
41	140	6.2	1127	2 T28317	ORF MSV156 hypotet
42	140	6.2	1712	2 C71618	hypothetical prote
43	139.5	6.1	1233	2 T40059	chromosome segrega
44	139.5	6.1	1979	2 C71622	hypothetical prote
45	139	6.1	1188	2 A71621	protein with 5'-3'

ALIGNMENTS

RESULT 1

A54498

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum

C>Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000

C:Accession: A54498

R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V.
Mol. Biochem. Parasitol. 27, 291-302, 1988

A:Title: Variation in the precursor to the major merozoite surface antigens of Plasmo

A:Reference number: A54498; MUID:88142999; PMID:2449612

A:Accession: A54498

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1701 <PEP>

A:Cross-references: GB:M19143; NID:G160412; PIDN:AAA29653.1; PID:G160413

C:Superfamily: major merozoite surface antigen

C:Keywords: surface antigen

Query Match

Best Local Similarity 85.6%; Score 1946.5; DB 2; Length 1701;

Matches 374; Conservative 7; Mismatches 8; Indels 5; Gaps 1;

QY	38	TDDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVITLKLPLAGVYSLKKQIEKNI	97
DB	1293	SEDNDEYLQQVVTGE-----AISVTMDNILSGFENEYDVITLKLPLAGVYSLKKQIEKNI	1347
QY	98	FTFNLMLNDILNSRLKRRKYFLDVLESQKNIIESSPKLLNSEQKNTLLKS	157
DB	1348	ITFNLMLNDILNSRLKRRKYFLDVLESQKNIIESSPKLLNSEQKNTLLKS	1407
QY	158	YKIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEKEKFPSSPTTSPSPAK	217
DB	1408	YKIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEKEKFPSSPTTSPSPAK	1467
QY	218	TBQKESKFLPPLTNIETLYNNLVNKIDDLINLAKINDCNVDEKDEAHVKTIKLSDLK	277
DB	1468	TBQKESKFLPPLTNIETLYNNLVNKIDDLINLAKINDCNVDEKDEAHVKTIKLSDLK	1527
QY	278	AIDDKIDLKPNPVDFBAIKKLINDDTKKMDLGLLSTGLVQNFPTNTIISKIEGKFDMDL	337
DB	1528	AIDDKIDLKPNPVDFBAIKKLINDDTKKMDLGLLSTGLVQNFPTNTIISKIEGKFDMDL	1587
QY	338	NISQHCVKQCCPNSGCFRHLDERECCLLANYKQEGDKCVENPNTCTNENNGCCDADA	397
DB	1588	NISQHCVKQCCPNSGCFRHLDERECCLLANYKQEGDKCVENPNTCTNENNGCCDADA	1647
QY	398	TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS	431
DB	1648	TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS	1681

RESULT 2

SAZQGM

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (strain RO-71)
N:Alternate names: 195K glycoprotein
C:Species: Plasmodium falciparum
C:Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000
C:Accession: A23386; S06361
R:Weber, J.L.; Leininger, W.M.; Lyon, J.A.
Nucleic Acids Res. 14, 3311-3323, 1986
A:Title: Variation in the gene encoding a major merozoite surface antigen of the human merozoite
A:Reference number: A23386; MUID:86205236; PMID:3517809
A:Accession: A23386
A:Molecule type: DNA
A:Residues: 1-1104 <WEB>
A:Cross-references: EMBL:X03831
R:Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Wolff, R.
Nucleic Acids Res. 16, 1206, 1988
A:Title: Merozoite surface protein sequence from the Camp strain of the human malaria parasite
A:Reference number: S06361; MUID:88143999; PMID:3278296
A:Accession: S06361
A:Molecule type: DNA
A:Residues: 1104-1726 <WEB2>
A:Cross-references: EMBL:X03831
C:Comment: The merozoite stages of different strains have strain-specific surface antigen sequences.
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat
F:1-19/Domain: signal sequence #status predicted <SIG>
F:17-76/Region: 3-residue repeats (S-G-T)
F:67-87, 91-96, 100-105, 109-120/Region: 3-residue repeats (S-G-T)
F:757-765/Region: 3-residue repeats (T-E-E)
F:133, 272, 501, 567, 638, 827, 839, 924, 944, 990, 1016, 1114, 1221, 1613, 1658/Binding site: carboxy

Query Match 85.1%; Score 1935.5; DB 1; Length 1726;
Best Local Similarity 94.4%; Pred. No. 1.8e-91;
Matches 372; Conservative 8; Mismatches 9; Indels 5; Gaps 1;

Qy 38 TDDDDKAMADIGSTEGRTMAISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQIEKNI 97
Db 1318 SEDNDEYLDQVVTGE-----AISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQIEKNI 1372
Qy 98 FTFNLNLDILNSRLKRRKYPFLDVLSDLMQFKHISNEYIIDSFKLLNSEQNTLLKS 157
Db 1373 FTFNLNLDILNSRLKRRKYPFLDVLSDLMQFKHISNEYIIDSFKLLNSEQNTLLKS 1432
Qy 158 YKIKESVENDIKFAQEGISYEVYKAKYKDDLESIKKVIKEEKEKFPSPPTPPSPAK 217
Db 1433 YKIKESVENDIKFAQEGISYEVYKAKYKDDLESIKKVIKEEKEKFPSPPTPPSPAK 1492
Qy 218 TDEQKESKFLPFTLTNETLNNLVNKIDDDYILNKLKAKINDCNVEKDEAHVKITKLSDLK 277
Db 1493 TDEQKESKFLPFTLTNETLNNLVNKIDDDYILNKLKAKINDCNVEKDEAHVKITKLSDLK 1552
Qy 278 AIDDKIDLFPKNYPDFEATKLLINDTKMDLGLKLLSTGLVQNFNTIISKLIIEGKFQDML 337
Db 1553 AIDDKIDLFPKNHDFEATKLLINDTKMDLGLKLLSTGLVQNFNTIISKLIIEGKFQDML 1612
Qy 338 NISQHCVKKQCPNSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGGCCADA 397
Db 1613 NISQHCVKKQCPNSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGGCCADA 1672
Qy 398 TCTEEDSGSSRRKTKTCECTKPDSPYPLFDGIFCSS 431
Db 1673 KCTEEDSGSNKKTKTCECTKPDSPYPLFDGIFCSS 1706

RESULT 3

A26868
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (strain RO-71)
C:Species: Plasmodium falciparum
C:Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000
C:Accession: A26868
R:Tanabe, K.; Mackay, M.; Goman, M.; Scalfe, J.G.
J. Mol. Biol. 195, 273-287, 1987

A:Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum
A:Reference number: A26868; MUID:88011243; PMID:3079521
A:Accession: A26868
A:Molecule type: DNA
A:Residues: 1-1701 <TAN>
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1701/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 84.6%; Score 1925.5; DB 2; Length 1701;
Best Local Similarity 94.2%; Pred. No. 5.8e-91;
Matches 371; Conservative 7; Mismatches 11; Indels 5; Gaps 1;

Qy 38 TDDDDKAMADIGSTEGRTMAISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQIEKNI 97
Db 1293 SEDNDEYLDQVVTGE-----AISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQIEKNI 1347
Qy 98 FTFNLNLDILNSRLKRRKYPFLDVLSDLMQFKHISNEYIIDSFKLLNSEQNTLLKS 157
Db 1348 FTFNLNLDILNSRLKRRKYPFLDVLSDLMQFKHISNEYIIDSFKLLNSEQNTLLKS 1407
Qy 158 YKIKESVENDIKFAQEGISYEVYKAKYKDDLESIKKVIKEEKEKFPSPPTPPSPAK 217
Db 1408 YKIKESVENDIKFAQEGISYEVYKAKYKDDLESIKKVIKEEKEKFPSPPTPPSPAK 1467
Qy 218 TDEQKESKFLPFTLTNETLNNLVNKIDDDYILNKLKAKINDCNVEKDEAHVKITKLSDLK 277
Db 1468 TDEQKESKFLPFTLTNETLNNLVNKIDDDYILNKLKAKINDCNVEKDEAHVKITKLSDLK 1527
Qy 278 AIDDKIDLFPKNYPDFEATKLLINDTKMDLGLKLLSTGLVQNFNTIISKLIIEGKFQDML 337
Db 1528 AIDDKIDLFPKNHDFEATKLLINDTKMDLGLKLLSTGLVQNFNTIISKLIIEGKFQDML 1587
Qy 338 NISQHCVKKQCPNSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGGCCADA 397
Db 1588 NISQHCVKKQCPNSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGGCCADA 1647
Qy 398 TCTEEDSGSSRRKTKTCECTKPDSPYPLFDGIFCSS 431
Db 1648 TCTEEDSGSSRRKTKTCECTKPDSPYPLFDGIFCSS 1681

RESULT 4

S47282
merozoite surface antigen 1 - malaria parasite (Plasmodium falciparum) (strain RO-71)
C:Species: Plasmodium falciparum
A:Variety: strain RO-71
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 09-Jun-2000
C:Accession: S47282
R:Tolle, R.; Bujard, H.; Cooper, J.A.
submitted to the EMBL Data Library, July 1994
A:Description: Plasmodium falciparum: recombination within the C-terminal region of merozoite surface antigen
A:Reference number: S47282
A:Accession: S47282
A:Molecule type: DNA
A:Residues: 1-651 <TOL>
A:Cross-references: EMBL:Z35329; NID:g535257; PIDN:CAA84558.1; PID:g535258
A:Experimental source: strain RO-71
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen

Query Match 84.3%; Score 1917; DB 2; Length 651;
Best Local Similarity 93.9%; Pred. No. 5.3e-91;
Matches 370; Conservative 9; Mismatches 9; Indels 6; Gaps 2;

Qy 38 TDDDDKAMADIGSTEGRTMAISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQIEKNI 97
Db 244 SEDNDEYLDQVVTGE-----AISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQIEKNI 298
Qy 98 FTFNLNLDILNSRLKRRKYPFLDVLSDLMQFKHISNEYIIDSFKLLNSEQNTLLKS 157
Db 299 FTFNLNLDILNSRLKRRKYPFLDVLSDLMQFKHISNEYIIDSFKLLNSEQNTLLKS 358

QY 158 YKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKEFPSSPTTPSPAK 217
|||||
Db 359 YKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKE-FPSSPTTPSPAK 417
QY 218 TDEQKESKFLPLNIETLYNNLVNKIDYILINLAKAKINDCNVDEAHVKITKLSDLK 277
Db 418 TDEQKESKFLPLNIETLYNNLVNKIDYILINLAKAKINDCNVDEAHVKITKLSDLK 477
QY 278 ATDDKIDLPKPNYPDEAIKKLNDOTKKDMLGKLLSTGLVQNPNTIISKLEGGFQDML 337
Db 478 ATDDKIDLPKPNYPDEAIKKLNDOTKKDMLGKLLSTGLVQNPNTIISKLEGGFQDML 537
QY 338 NISQHCVKKQCPENSGCFRHLDERECKCLLNKYQEGDKVCVENPNPTCNENNGGCDADA 397
Db 538 NISQHCVKKQCPENSGCFRHLDERECKCLLNKYQEGDKVCVENPNPTCNENNGGCDADA 597
QY 398 TCTEEDSGSSRKITCECTKPDSPYPLFDGIFCSS 431
Db 598 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSS 631

RESULT 5
A:5948
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
C:Species: Plasmodium falciparum
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jun-2000
C:Accession: A45948
R:Chang, S.P.; Kramer, K.J.; Yanaga, K.M.; Kato, A.; Case, S.E.; Siddiqui, W.A.
Exp. Parasitol. 67, 1-11, 1988
A:Title: Plasmodium falciparum: gene structure and hydropathy profile of the major merozo
A:Reference number: A45948; MUID:89005525; PMID:3049134
A:Accession: A45948
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1726 <CHA>
A:Cross-references: GB:M37213
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 84.2%; Score 1916.5; DB 2; Length 1726;
Best Local Similarity 93.7%; Pred. No. 1:7e-90;
Matches 369; Conservative 8; Mismatches 12; Indels 5; Gaps 1;
QY 38 TDDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVIYKPLAGVYRSKQKEKNI 97
Db 1318 SEDNEDYLDQVVTGE-----AISVTMDNILSGFENEYDVIYKPLAGVYRSKQKEKNI 1372
QY 98 FTFNLNLDILNSRLKRRKYFDLVLESQKFKHSSNEYIETDSFKLLNSQKNTLLKS 157
Db 1373 FTFNLNLDILNSRLKRRKYFDLVLESQKFKHSSNEYIETDSFKLLNSQKNTLLKS 1432

QY 158 YKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKEFPSSPTTPSPAK 217
Db 1433 YKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKEFPSSPTTPSPVK 1492
QY 218 TDEQKESKFLPLNIETLYNNLVNKIDYILINLAKAKINDCNVDEAHVKITKLSDLK 277
Db 1493 TDEQKESKFLPLNIETLYNNLVNKDDYILINLAKAKINDCNVDEAHVKITKLSDLK 1552
QY 278 ATDDKIDLPKPNYPDEAIKKLNDOTKKDMLGKLLSTGLVQNPNTIISKLEGGFQDML 337
Db 1553 ATDDKIDLPKPNYPDEAIKKLNDOTKKDMLGKLLSTGLVQNPNTIISKLEGGFQDML 1612
QY 338 NISQHCVKKQCPENSGCFRHLDERECKCLLNKYQEGDKVCVENPNPTCNENNGGCDADA 397
Db 1613 NISQHCVKKQCPENSGCFRHLDERECKCLLNKYQEGDKVCVENPNPTCNENNGGCDADA 1672
QY 398 TCTEEDSGSSRKITCECTKPDSPYPLFDGIFCSS 431
Db 1673 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSS 1706

RESULT 6
A:5545
major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (fragments
C:Species: Plasmodium falciparum
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
C:Accession: A45545
R:Blackman, M.J.; Ling, I.T.; Nicholls, S.C.; Holder, A.A.
Mol. Biochem. Parasitol. 49, 29-33, 1991
A:Title: Proteolytic processing of the Plasmodium falciparum merozoite surface protei
A:Reference number: A45545; MUID:92131048; PMID:1775158
A:Accession: A45545
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-400 <BLA>
A:Note: sequence extracted from NCBI backbone (NCBIN:77612, NCBIP:77621)
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

Query Match 51.6%; Score 1175; DB 2; Length 400;
Best Local Similarity 56.6%; Pred. No. 2.8e-53;
Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;
QY 34 PDLGTDGDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVIYKPLAGVYRSKQKI 93
Db 5 PIFGESEED--YDDLQGVVTGEAVTPSV-IDNLTSLKIENEYEVLYKPLAGVYRSKQK 61
QY 94 EKNIFTNLDILNSRLKRRKYFDLVLESQKFKHSSNEYIETDSFKLLNSQKNT 153
Db 62 ENNVTFNVKVDILNSRKNRKNFRNVLSDIPIKDTSSNVYKVDPIKFLNKEKRD 121
QY 154 LKSYKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKEFPSSPTTP 213
Db 122 FLSSYVINKSDITDINFANDVLGYKYLSEKYKSDLSIKKYI----- 165
QY 214 SPAKTDEQKESKFLPLNIETLYNNLVNKIDYILINLAKAKINDCNVDEAHVKITKL 273
Db 166 ----NDKOGENEKYLPLNNIETLYKTVANDKIDLFVHLEAKVLYNTEKSNVEYVAKEL 221
QY 274 SLDKATDDKIDLPKPNYPDEAIKKLNDOTKKDMLGKLLSTGLV-QNFPNTIISKLEGG 332
Db 222 NYLKTQDKLADFKNNNFVGIADLSTDYNNHNLTKFLSTGWFENLAKTVLSNLDGN 281
QY 333 FODMLNISQHCVKKQCPENSGCFRHLDERECKCLLNKYQEGDKVCVENPNPTCNENNGG 392
Db 282 LQCLMLNISQHCVKKQCPENSGCFRHLDERECKCLLNKYQEGDKVCVENPNPTCNENNGG 341
QY 393 CDADACTEEDSGSSRKITCECTKPDSPYPLFDGIFCSS 431
Db 342 CDADAKTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSS 380

RESULT 7
S05603
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N:Alternate names: gp195 surface antigen
C:Species: Plasmodium falciparum
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
C:Accession: S05603; S04850
R:Myler, P.J.
submitted to the EMBL Data Library, April 1989
A:Reference number: S05603
A:Accession: S05603
A:Molecule type: mRNA
A:Residues: 1-1639 <MYL>
A:Cross-references: EMBL:X15063; NID:g9896; PIDN:CAA33163.1; PID:g9897
Nucleic Acids Res. 17, 5401, 1989
A:Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from pl
A:Reference number: S04850; MUID:89345116; PMID:2668887
A:Accession: S04850
A:Molecule type: mRNA
A:Residues: 1504-1639 <MYL2>
A:Cross-references: EMBL:X15063

C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1639/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 51.6%; Score 1175; DB 2; Length 1639;
Best Local Similarity 56.6%; Pred. No. 1.4e-52;
Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;

QY 34 PDLTDDDDKAMADIGSIEGRGTMAISVTMDNLSGFENEYDVLYLKPLAGVYRSLAKQI 93
Db 1244 PIGESED--YDDLQGVVTGEAVTPSV-IDNLSKTIENEYEVLYLKPLAGVYRSLAKQI 1300
QY 94 EKNIFTNMLNDILNSRLKRRKRYFLDVLDSLMQFHHISSNEYIIEEDSFKLLNSEQKNT 153
Db 1301 ENNVMTENVVVKDILNSRKNFKNVLESLLIPYKDLTSSNVVVKDPYKFLNKEKRDK 1360
QY 154 LLKSYKIKESVENDIKFAQEGISYYEKLAKYKDDLESIKKVIKEKEKFPSPPTTPP 213
Db 1361 FLSSYNYKDSIDTDFINFANDVLGYKILSEKYSKSDLSIKKYI----- 1404
QY 214 SPAKTDOKKESKFLPFLTNIETLYNNLVNKIDYTLNKLAKINDCNVKEDEAHVKITKL 273
Db 1405 ----NDKQGENEYKLPFLNIETLYNNLVNKIDYTLNKLAKINDCNVKEDEAHVKITKL 1460
QY 274 SOLKAIDDKIDLFPKNPYDFFAIKKLINDDTKKOMGLKLLSTGLV-QNFPNTIISKLEGG 332
Db 1461 NYLKTIOKDLADFKNNFVGIADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLDGN 1520
QY 333 FQDMLNISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNTCENNGG 392
Db 1521 LQGLMLNISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNTCENNGG 1580
QY 393 CDADATCTEEDSGSRKKTCTCTKPDSPYPLFDGIFCSS 431
Db 1581 CDADAKCTEEDSGSGKKTCTCTKPDSPYPLFDGIFCSS 1619

RESULT 8
SAZQK1
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
C:Species: Plasmodium falciparum
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 09-Jun-2000
C:Accession: A25120
R:Kackay, M.; Goman, M.; Hyde, J.E.; Scaife, J.; Certa, U.; Stunnenberg, H.; H
EMBO J. 4, 3823-3829, 1985
A:Title: Polymorphism of the precursor for the major surface antigens of Plasmodium fal
A:Reference number: A91030; MUID:86136024; PMID:3004972
A:Accession: A25120
A:Molecule type: DNA
A:Residues: 1-1631 <MAC>
C:Comment: The merozoite stages of different strains have strain-specific surface antige
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane prote
F:1-1631/Product: major merozoite surface antigen #status predicted <SIG>
F:20-1631/Region: 3-residue repeats (S-G-T/P)
F:1614-1631/Domain: membrane anchor #status predicted <MBN>
F:97,259,755,759,835,911,955,1049,1156,1165,1436,1563/Binding site: carbohydrate (Asn) (

Query Match 46.2%; Score 1051.5; DB 1; Length 1631;
Best Local Similarity 52.3%; Pred. No. 3e-46;
Matches 210; Conservative 62; Mismatches 103; Indels 25; Gaps 5;

QY 34 PDLTDDDDKAMADIGSIEGRGTMAISVTMDNLSGFENEYDVLYLKPLAGVYRSLAKQI 93
Db 1235 PIGESED--YDDLQGVVTGEAVTPSV-IDNLSKTIENEYEVLYLKPLAGVYRSLAKQI 1291
QY 94 EKNIFTNMLNDILNSRLKRRKRYFLDVLDSLMQFHHISSNEYIIEEDSFKLLNSEQKNT 153
Db 1292 ENNVMTENVVVKDILNSRKNFKNVLESLLIPYKDLTSSNVVVKDPYKFLNKEKRDK 1351

QY 154 LLKSYKIKESVENDIKFAQEGISYYEKLAKYKDDLESIKKVIKEKEKFPSPPTTPP 213
Db 1352 FLSSYNYKDSIDTDFINFANDVLGYKILSEKYSKSDLSIKKYI----- 1395
QY 214 SPAKTDOKKESKFLPFLTNIETLYNNLVNKIDYTLNKLAKINDCNVKEDEAHVKITKL 273
Db 1396 ----NDKQGENEYKLPFLNIETLYNNLVNKIDYTLNKLAKINDCNVKEDEAHVKITKL 1451
QY 274 SDLAIDDKIDLFPKNPYDFFAIKKLINDDTKKOMGLKLLSTGLV-QNFPNTIISKLEGG 332
Db 1452 IYLTIOKDLADFKNNFVGIADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLDGN 1511
QY 333 FQDMLNISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNTCENNGG 391
Db 1512 LARYVKHFTTPMRKKTMIQSSGCFRHLDERECKCLLNYKQEGDKCVENPNTCENNGG 1571
QY 392 CDADATCTEEDSGSRKKTCTCTKPDSPYPLFDGIFCSS 431
Db 1572 GCDADAKCTEEDSGSGKKTCTCTKPDSPYPLFDGIFCSS 1611

RESULT 9
A24594
probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodi
C:Species: Plasmodium falciparum
C:Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jun-2000
C:Accession: A24594
R:Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholl
Nature 317, 270-273, 1985
A:Title: Primary structure of the precursor to the three major surface antigens of Pl
A:Reference number: A24594; MUID:86014355; PMID:2995820
A:Accession: A24594
A:Molecule type: DNA
A:Residues: 1-1640 <HOL>
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 44.3%; Score 1007.5; DB 2; Length 1640;
Best Local Similarity 50.9%; Pred. No. 5.3e-44;
Matches 202; Conservative 65; Mismatches 105; Indels 25; Gaps 5;

QY 37 GTDDDDKAMADIGSIEGRGTMAISVTMDNLSGFENEYDVLYLKPLAGVYRSLAKQI 96
Db 1247 GESEED--YDDLQGVVTGEAVTPSV-IDNLSKTIENEYEVLYLKPLAGVYRSLAKQI 1303
QY 97 IFTFNLNDILNSRLKRRKRYFLDVLDSLMQFHHISSNEYIIEEDSFKLLNSEQKNTLLK 156
Db 1304 VMTFNVNVKDIILNSRKNFKNVLESLLIPYKDLTSSNVVVKDPYKFLNKEKRDK 1363
QY 157 SYKIKESVENDIKFAQEGISYYEKLAKYKDDLESIKKVIKEKEKFPSPPTTPPSPA 216
Db 1364 SYNYKDSIDTDFINFANDVLGYKILSEKYSKSDLSIKKYI----- 1404
QY 217 KTDEOKESKFLPFLTNIETLYNNLVNKIDYTLNKLAKINDCNVKEDEAHVKITKLSDL 276
Db 1405 -NOKQGENEYKLPFLNIETLYNNLVNKIDYTLNKLAKINDCNVKEDEAHVKITKLSDL 1463
QY 277 KAIDDKIDLFPKNPYDFFAIKKLINDDTKKOMGLKLLSTGLV-QNFPNTIISKLEGG-Q 334
Db 1464 KTIOKDLADFKNNFVGIADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLDGN 1523
QY 335 DMLNISQHCYKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNTCENNGGCD 394
Db 1524 VYKHFTTPMRKKTMIQSSGCFRHLDERECKCLLNYKQEGDKCVENPNTCENNGGCD 1583
QY 395 ADATCTEEDSGSRKKTCTCTKPDSPYPLFDGIFCSS 431
Db 1584 ADATCTEEDSGSGKKTCTCTKPDSPYPLFDGIFCSS 1620

RESULT 10
A39401
merozoite surface antigen 1 precursor - Plasmodium vivax

C:Species: Plasmodium vivax
C:Date: 28-Feb-1992 #sequence_revision 28-Feb-1992 #text_change 24-Nov-1999
C:Accession: A39401
R:del Portillo, H.A.; Longacre, S.; Khouri, E.; David, P.H.
Proc. Natl. Acad. Sci. U.S.A. 88, 4030-4034, 1991
A:Title: Primary structure of the merozoite surface antigen 1 of Plasmodium vivax reveal
A:Reference number: A39401; MUID:91219506; PMID:2023952
A:Accession: A39401
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1726
A:Cross-references: GB:M60807
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 35.9%; Score 816; DB 2; Length 1726;
Best Local Similarity 39.8%; Pred. No. 3.5e-34;
Matches 177; Conservative 87; Mismatches 141; Indels 40; Gaps 10;

QY 8 SGLVPRGCMETAAKFERQHM-----SPDLGTDGDDKAMADIGSIEGRGT 57
DB 1280 TGVVPP--GVVESAQAQKAQDAEDYDKVIELPFGNDDGDEEDQVTTGE----- 1331
QY 58 ATSVTMDNLSCFENEYDVIYKPLAGVYRSLLKQIEKNIFTNMLNLDLNSRLKRY 117
DB 1332 ASEAPEILVPAGISDYDVVYKPLAGMYKKIKKQLENHVNAPNTNITDMLDSRLKRY 1391
QY 118 FLDVLESLLMQPKHSSNEYIIESDFKLLNSQKNTLLKSYKIKESVENDIKFAQEGIS 177
DB 1392 FLEVNSDLNPKYSPSGEYIIPKYLKLDLEKKLLKLSYKIGASIDKDLCTANDGVN 1451
QY 178 YEKVLAKYKDDL-----ESIKV---IKEEKEKFPSSPTTPPSPAKTDEQKES---KF 227
DB 1452 YNKGKELYKTHLTAVNEEVKKVEADIKAEEDDKIKKIGSDSTKTEKTSMAKAELEKY 1511
QY 228 LPLFTNIETLYNVLNKKIDYLLINKAKINDCNVEKDEAHVKITKLSLDKADDKLDLEK 287
DB 1512 LPLNSLQKEYESLSKVNTYDNLKVINCOLEKKEAEITVKKLDQYKNKMDERLEEK 1571
QY 288 NPYDFEAIKLLINDDTKMDLGLKLLSTGLV-QNFPNTIISKLTIEGKFODMLNISQHCYK 346
DB 1572 -----KSEKK--NEVKSGLLEKLMKSLIKENESKILLSQLLNVTQTLTWSSEHTCID 1624
QY 347 KOCPSNGCFRHLDBRECKCLLNKQEGDKCVENPNTCNENNGGCCDADATCTEDSGS 406
DB 1625 TNVPDAACRYLDGMEWEERCLLTFAEKGKCVPGSNVTCKDNNGGCAPEAECKMTDS-- 1682
QY 407 SRKKITCECTKPDSPYLPFGIFCSS 431
DB 1683 --NKIVCKCTKEGSEPLFEGVFCSS 1705

RESULT 11
A45604
major blood-stage surface antigen Pv200 - Plasmodium vivax
C:Species: Plasmodium vivax
C:Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999
C:Accession: A45604
R:Gibson, H.L.; Tucker, J.E.; Kaslow, D.C.; Krettli, A.U.; Collins, W.E.; Kieffer, M.C.;
Mol. Biochem. Parasitol. 50, 325-333, 1992
A:Title: Structure and expression of the gene for Pv200, a major blood-stage surface antigen
A:Reference number: A45604; MUID:92158013; PMID:1371329
A:Accession: A45604
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1751 <GIB>
A:Cross-references: GB:M75674; NID:g160608; PID:g457336
A:Note: sequence extracted from NCBI backbone (NCBIN:83591, NCBIP:83592)
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 34.9%; Score 793.5; DB 2; Length 1751;
Best Local Similarity 38.1%; Pred. No. 5.1e-33;

Matches 174; Conservative 87; Mismatches 135; Indels 61; Gaps 10;

QY 9 SGLVP-RGSG-----MKETAARFERQHM-----SPDLGTDGDDKAMADIGSIEGRGT 56
DB 1301 AGSVPTATGGAAGSTEENVAKAQDAEDYDKVIALPFLGNNDDGDEEDQVTTGE----- 1356
QY 57 MAISVTMDNLSCFENEYDVIYKPLAGVYRSLLKQIEKNIFTNMLNLDLNSRLKRY 116
DB 1357 -AESEAPEILVPAGISDYDVVYKPLAGMYKTIKKOLENHVNAPNTNITDMLDSRLKRY 1415
QY 117 YFLDVLESLLMQPKHSSNEYIIESDFKLLNSQKNTLLKSYKIKESVENDIKFAQEGI 176
DB 1416 YFLEVNSDLNPKYSSSGEYIIPKYLKLDLEKKLLKLSYKIGASIDKDLCTANDGV 1475
QY 177 SYEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTPPSPAKTDEQKES----- 224
DB 1476 TYVKNKGELYKTHLQGVKTEIKKVEDDI-----KKODELKLKLVNSQDSKK 1524
QY 225 -----SKFLPFTNIETLYNVLNKKIDYLLINKAKINDCNVEKDEAHVKITKLS 275
DB 1525 EFTAKAELEKYLPLFNSLQKEYESLSKVNTYDNLKVINCOLEKKEAEITVKKLD 1584
QY 276 LKAIDDKIDLEKNPYDFAIKKLLINDTKKMDLGLKLLSTGLV-QNFPNTIISKLTIEGKF 334
DB 1585 YNKMDEKLEYK-----KSEKK--NEVKSGLLEKLMKSLIKENESKILLSQLLNVTQ 1637
QY 335 DMLNISQHCYKQCCPNSGCFRHLDBRECKCLLNKQEGDKCVENPNTCNENNGGCC 394
DB 1638 LLTMSSEHTCIDTNVPDAACRYLDGTEWRCLLTFAEKGKCVPGSNVTCKDNNGGCA 1697
QY 395 ADATCTEEDSGSRRKKTCECTKPDSPYLPFGIFCSS 431
DB 1698 PEAECKMTDS---NKIVCKCTKEGSEPLFEGVFCSS 1730

RESULT 12
A28121
major merozoite surface antigen - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 26-Aug-1999
C:Accession: A28121
R:Burns Jr., J.M.; Daly, T.M.; Vaidya, A.B.; Long, C.A.
Proc. Natl. Acad. Sci. U.S.A. 85, 602-606, 1988
A:Title: The 3' portion of the gene for a Plasmodium yoelii merozoite surface antigen
A:Reference number: A28121; MUID:88124889; PMID:2448778
A:Accession: A28121
A:Molecule type: DNA
A:Residues: 1-680 <BUR>
A:Cross-references: GB:J03612; NID:g160678; PID:g160679
A:Experimental source: strain 17XL
A:Note: the authors translated the codon GTA for residue 429 as Leu
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 24.5%; Score 556.5; DB 2; Length 680;
Best Local Similarity 31.5%; Pred. No. 2.3e-21;
Matches 135; Conservative 86; Mismatches 164; Indels 43; Gaps 11;

QY 10 GLVPRGSGMKETAARFERQHMSPDLGTDGDDKAMADIGSIEGRGTMAISVTMDNLSC 69
DB 270 GAVP-MSGTDRVAG-----SSVDNEDDDIYQIAS-GQSE-----DAPEKDLISE 313
QY 70 FENEYDVIYKPLAGVYRSLLKQIEKNIFTNMLNLDLNSRLKRYFLDVLSDLMQF 129
DB 314 FTTNESLYVYTKRLGTYKSLKHHMLREFSTIKEDMTNGLNKKSKQRNDFLEVLSHELD 373
QY 130 KHTSSNEYIIESDFKLLNSQKNTLLKSYKIKESVENDIKFAQEGISYEKVLAKYKDD 189
DB 374 KDLSTNKYIRNPYQLDNDKDKQIVNLKYATKGINEDIETTTDGTGKFFKNKVEYNTQ 433
QY 190 LESIKKVIKEEKEKFPSSPTTPPSPAKTDEQKES--SKFLPFTNIETLYNVLNKKIDD 247
DB 434 LAAPKQIATIE-----AETNDTKEEKKKYIPILEDLKLGYETVIGQAE 479

Db 1730 KISLYRNNIETSIENEYNSFSKSIKSDGLLEDEL--KSLKHSTSEIETIKSGLQEQI 1786

QY 202 EKFPSPPTTSPAKTDQKESKFLPFLTNIETLYNNLVNKIDDYLLINLAKINDCNV 261
|||
Db 1787 DXFEV-----EFKNHK-----BLLEVDNNILESKILNCV 1820
QY 262 EKDEAHVKITKLSLKAIDDKIDL---FKNPYD-----FEATKKLINDTK----- 304
|||
Db 1821 QFNKF---LSEIKD-NLVYKSDLRAEFEDYDKINFQENIOENPNFKLUDSELEKNNSTF 1876
QY 305 -----KDLGKLLST-----GLVQNPNTIISKLEGFQDMLNISQHCQVKKOCPPENS 353
|||
Db 1877 LEAYSLSKDLKLEKWLTKNEIGLAQYKNN-----FENVNKEFYNIQK-----ETL 1922
QY 354 GCFRHLDEREECKLLNYKQEGDKCVEN 381
|||
Db 1923 GIIEIFNE-----LKLQESIKSIKN 1943
RESULT 18
A64505
P115 homolog - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 02-Jun-2000
C:Accession: A64505
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
raon, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Bordovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: A64505
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1169 <BUL>
A:Cross-references: GB:U67604; GB:L77117; NID:g1592224; PID:g1500543; TIGR:MJ1643
A:Map position: FOR1623481-1626990
C:Genetics:
C:Superfamily: chromosome segregation protein SMC1

Query Match 7.1%; Score 160.5; DB 2; Length 1169;
Best Local Similarity 24.2%; Pred. No. 0.79;
Matches 86; Conservative 59; Mismatches 110; Indels 101; Gaps 17;

QY 94 EKNIFENLNLNLSRLK---KRYFLDVLESFLMFKHSSNEYIIESFKLL----- 146
|||
Db 85 ENNAFVNADKVGIUR-RIKSGSETDYLVWKNDEKRRKMTKHEII--DLFRRLGLLG 141
QY 147 -NSEQKNTLLK-----SYKIKESVENDIKFAQEGISYVEKVLAK 185
|||
Db 142 DNVISQDGLLKIINISPIERRKIIDEISGIAEFDEKKKAEELKARELIEMIDIRISE 201
QY 186 YKDDLESKKVKEKEKFPSPPTTSPAKTDQKESKFLPFL----- 231
|||
Db 202 VENNLKLLKE--KDAERY-----IKLNEELAAKAYALIKKVSVLNVLLENIQ 249
QY 232 ---TWIETLYNNLVNK---IDYLLINLAKINDC-----NVEKDEAHVKITKL----- 273
|||
Db 250 NDIKLEELKNEFLSKVREIDVEIENLRLNIINELNKGNEEVELEHKSIRKELEVEI 309
QY 274 -SDLKAIDDKI-DLPKNPYDFEAIKKLLINDTKKMLGKLLSTGLVQNPNTIIS---K 327
|||
Db 310 ENDKKVLDSINELKVEIENKKEIKETQKK-----IENRDSIIEKEQOIK 359
QY 328 LIEGKFQDMLNISQHCQVKKOCPPENSFCFRHLDEREECKLLNYKQEGDKCVENP 383
|||
Db 360 EIEEKIKN-LNYKER-LKEATAESESIIKHLKESE-----MEIADETAKNQ 405

RESULT 19

T18372

repeat organellar protein - Plasmodium chabaudi

C:Species: Plasmodium chabaudi

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18372
R:Werner, E.B.; Taylor, W.R.; Holder, A.A.
Mol. Biochem. Parasitol. 94, 185-196, 1998
A:Title: A Plasmodium chabaudi protein contains a repetitive region with a predicted
A:Reference number: Z18922; MUID:98418765; PMID:9747969
A:Accession: T18372
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1939 <WER>
A:Cross-references: EMBL:U43145; NID:g1151157; PID:g1151158; PIDN:AAC63403.1

Query Match 6.9%; Score 156.5; DB 2; Length 1939;
Best Local Similarity 22.5%; Pred. No. 2.3;
Matches 97; Conservative 57; Mismatches 141; Indels 137; Gaps 18;

QY 71 ENEDYVILKPLAGVYRSLLKKOIE-----KNITFFNL--NLNDI--- 107
|||
Db 114 EKQINKYEKELNEKSEFIKROMELLKELNINKENKINKEITTLKREELNDIESE 173
QY 108 -----LNSRLKRRKYFLDVLESFLMFKHSSNEYIIESFKLLNS---EOKNTL 154
|||
Db 174 YTEKNKEKELNVEYTNKMSLDKLTCEVQEKK-----DNLEKINKKVIKENNL 223
QY 155 LKSYKIVKESVE-----NDIKFAQD--GISYVEK----- 181
|||
Db 224 RELKEFMKEKNEIIESLDGTINDKKNAYEKLSEIFEKRRKMIEMLSDKLIEKEENFANKQ 283
QY 182 -----VLAKYKDDLESIKKVKKEKFPSPPTTSPAKTDQKESKFLPFLT 232
|||
Db 284 AKLEKENETIIEKLL-DIESREKDFKSEKKEKFAFM--ENELNLSKSLDNACOMEVYKL 340
QY 233 NIETLYNNLV-----NKKIDDYLLINLAK---INDCNVEKDEAH---VKITKL--- 273
|||
Db 341 EIKDLSQSLVEKEREFEIKNEVDKDKINNMKEKLSINDKGDINTVLHSEEEKINKLLKE 400
QY 274 --SDLKAIDDKIDLKPNPYDFEAIKKLLINDTKKMLGKLLSTGLVQNPNTIISKLIEG 331
|||
Db 401 KETELNEIHKYNNL-----EITIKNELKEEKEEKKKAHTVEVTNLTKEI--KLEK 453
QY 332 FKQDMLNISQHCQVKKOCPPENSFCFRHLDEREECKLLNYKQEGDKCVENPNTNENNG 391
|||
Db 454 KTED-----AKEGHKNELNQLSKLNEKDN---INKNENTELNDKIS 495
QY 392 GCDADATCTED 403
|||
Db 496 SLNSEVNILNKD 507

RESULT 20

B70356

chromosome assembly protein homolog - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Jun-2000

C:Accession: B70356

R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: B70356

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1156 <AQF>

A:Cross-references: GB:AE000699; NID:g2983236; PIDN:AAC06839.1; PID:g2983243; GB:AE00

A:Experimental source: strain VF5

C:Genetics:

A:Gene: xcpC

C:Superfamily: chromosome segregation protein SMC1

Query Match

Best Local Similarity

Matches

91; Conservative

81; Mismatches

129; Indels

105; Gaps

21;

Qy	39	DDDDKAMADIGSTEGRTMAIS--VTMDNILSGF-----ENEVDVIY-----	78
Dd	96	EDEEVISRKVSDGRSIFKINGQVVRERDLKDFLAKAGIYETAVNVYQGDIVAFKLMT	155
Qy	79	-----LKPLAGV--YRSLKK-----QIEKNITFNLDNLN--SRLLKKRYFLD	120
Dd	156	PVERRKIIIEISGIGERYEKKEKALEAEVLKIKEIDILILEETSNOQLRKEKEKLE	215
Qy	121	VLEDLMQPFKHISSNEYIIEDSFKLNSBQKNTLLSKYIKIESVENDIKFAOEGISYYE	180
Dd	216	KFK-ELQRIKRETEAKILLKEREKLL--KERERILNELSSLSLE-DITFQ---IQENE	268
Qy	181	KVLAKYKDOLLESIKKIY---KEEKEFPSPPTTPPSPAKTDEQKESFLPFLNIETL	237
Dd	269	KELNERERLLKEYNEKIMPPFEKVGKFTABIEAERSIKERELKESP--NRVKNLEEL	326
Qy	238	YNLVNKIDDLYLNLAkakNDcNVE-----KDAHVKITKLSOLKAIDOKIDLfKNP	289
Dd	327	NNLLSDKE----NLREVGTVTLOLEKLEKEVKSUKEVEREKLEEEERLKI----	378
Qy	290	YDFEAIKKLINDOTKMDLGLSTGLVQ-----NFPNTI-----ISKLIIEKGFO	335
Dd	379	-TFDEVKKL--EEEKLEKTEKLSNKEQELEIQANLKNKTRIKEDINKLISEREK	435
Qy	336	MLNISQHQ-----CVKKQCPCPSCFRHLDERECECKLLNYKQE	374
Dd	436	IKEIKEQEQEIIRLKAIRKK-----EEEE---LRNLTOE	466
 RESULT 21 B97116			
Chromosome segregation SMC protein, ATPase [Imported] - Clostridium acetobutylicum			
C:Species: Clostridium acetobutylicum			
C:date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2003			
C:Accession: B97116			
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gilchrist, D.A.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.			
J. Bacteriol. 193, 4823-4838, 2001			
A:title: Genome Sequence and Comparative Analysis of the Solvent-Producing Strain ATCC8739			
A:Reference number: A96900; MUID:21359325; PMID:21359325			
A:Accession: B97116			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-1191 <KUP>			
A:cross-references: GB:AE001437; PIDN:AAK79717.1; PID:gl5024720; GSPDB:GN001437			
A:Experimental source: Clostridium acetobutylicum ATCC8739			
C:Genetics:			
A:Gene: CAC1751			

```

Qy 298 LINDTTK---DMLGKLLSTGLVON--FPTTIISKLIEGKFQDMLNISOHCVKKQCP 351
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 408 EIESNESLIKRAGEYSKSLKINEVTNLTSEELV--KINDKISGYENQ-IRENRSK 464
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Qy 352 NSGCFRHLDRREECKCLLNYSKQEGDKCVENPNTCN 387
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 465 ISKLNRIISDEEKNRELNSK--SNKLEANKNMLIN 498
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

RESULT 22
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28676, A45521
R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A:Title: Comparison of two members of a multigene family coding for high-molecular ma
A:Reference number: 220507; MUID:97077455; PMID:8920022
A:Accession: T28676
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AAB41263.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: A45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KEE>
A:Cross-references: GB:M34281

Query Match      6.7%; Score 152; DB 2; Length 2401;
Best Local Similarity 21.5%; Pred. No. 4.9;
Matches 91; Conservative 82; Mismatches 159; Indels 92; Gaps 23;

Qy 64 DNILSGFENEVDVYTLKPLAGVYRSCLKQIEKNFTFNMLNDILNSRLKRRKYFLDV-- 121
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 285 NEIYNTIKSDPKDIYVGDIOLYNEMFVSQVQESNIEHIEKNTEILTTLTKIDNVYNNIQN 344
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Qy 122 LESDLMQ--FKHISSNEYILE--DSFKLLNSEQKNTL--LKSYYKIKESVENDI-KFA 172
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 345 METETVSHLKNIEIETNNKLSLTIIDLIKIYGEITNELNKTLEDFKNKEKGLSNKIDEYA 404
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Qy 173 QEGIS---YYEKVL---AKYKD--DLESIK-KVIEKEKEFPSPPTTPSPAKTDSQKK 223
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 405 KENVOLNVYKSNILEIKKHYNDQINIDNIKEKAQNYDQFEHMKTIPTNEMKYQKPSI 464
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Qy 224 ESKEL--PFLTNIETLYNN-----LVNKI-----DDYLINLKAIND 258
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 465 EIKIMKDFLSKVNK-VNDFDKVYKEKVESEHNNKFTELTNKRIETVSDIEIKKYENKFD 523
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Qy 259 ---CNVEKDEAHVKITKSLDKAIDDKIDLFPKPYDFEALKKLINDTK-KDWLGKLLS 313
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 524 SKSLINETKKSIEBEYQNIPLFKKVDYIYKVLNTN--ELITNCHNQTTLKOKLNQNIK 581
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Qy 314 TGLVQNFPTTIISKLIEGKFQDM-----LNISOHCVKKQ-----CPEN 352
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 592 TIKETN----SIDKIYTDKEENILTDKKTELTFTGLSLNNHESNNKELLTYFYDLKAN 637
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Qy 353 SG-----CFRLDERECKCLLNYSKQEG--DKCVENPNTCNENNGCGDAATCTEED 403
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 638 LGKNKENMLYKQFNEKE--KAVEDIKKKNVNDINKIVSNIETIITYTSYININED---TENE 692
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Qy 404 SGSS 407
      |   |
Db 693 IGS 696

RESULT 23

```

B59102
hypothetical protein px01-90 - Bacillus anthracis virulence plasmid px01
C:Species: Bacillus anthracis
C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 11-May-2000
C:Accession: B59102
R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehler, J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of px01, the large Bacillus anthracis plasmid harbored by B. anthracis strain Sterne
A:Reference number: A59091; MUID:99445483; PMID:10515943
A:Accession: B59102
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-652 <OKI>
A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AA032394.1; PID:g4894306
A:Experimental source: strain Sterne
A:Note: similar to hypothetical protein; PFB0765w (980 aa); Plasmodium falciparum (AE001001)
C:Genetics:
A:Gene: px01-90
A:Genome: plasmid

Query Match 6.6%; Score 149.5; DB 2; Length 652;
Best Local Similarity 21.2%; Pred. No. 1.5;
Matches 83; Conservative 70; Mismatches 132; Indels 107; Gaps 15;

Qy 18 MKETAARAFERQHMDSPLDGTDDDDKAMADIGSIEGRGTMAISVTMDNLSGFP----- 71
Db 284 LKESQSRLELNKXDSNRLEINSEIKLND-----RKAELLSLIMELIKQOSEFKKIK 337
Qy 72 NEYDVI-----YKPLAGVYRSUKKQIEKNIFTFN--NLNDILNSLRKKRYFLDVLES 124
Db 338 NEKDLKKREDLNRAESKELAKKAEALNTKLVFLKVOEALNKKSGQVLYIINKLDN 397
Qy 125 DLMOFKHSSNEYIIEFSKLLNSEQNTLKSY--RYIK--ESVENDIRFAQEGISYYE 180
Db 398 ELRE-----LADKYK--NSDNKISRLKNHIGEYKNQLEKIELEPCNKKIDNTK 445
Qy 181 KVLK-----KYKDDLE-SIKKVIKEKEKFPSPSP 209
Db 446 KQAEFDKSNKQOESELSVLQNKKIDELGKRHKHROELESASQKALDEAKENKKLAE 505
Qy 210 TTPSPAKTDEOKESKFLPELTNIETLYNNLVNKIDYILINKAKI-----NDC 259
Db 506 KQPERHINDKLLK-----LNDRWLTNNELNRLNSELOKYNKTLGYSOEHOYNNAL 558
Qy 260 NVEKDEAHVKITLSDLAIDDK-----IDLKPNYPDEAIFKILINDDTKMDLGLK 311
Db 559 MKRETLSSKKVDEYQKQSEIDKKNYIELNROLDELSPKPKNEQVQLL-----KNQINEL 612
Qy 312 LSTGLVQNFNPNTIISKLECKFQDMLNISQHO 343
Db 613 -----NSSINQILNHRKNEINNITQRK 634

RESULT 24
C97325
hypothetical protein CAC3461 [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: C97325
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: C97325
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-909 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK81390.1; PID:g15026552; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3461

Query Match 6.5%; Score 147; DB 2; Length 909;
Best Local Similarity 20.8%; Pred. No. 2.9;
Matches 99; Conservative 80; Mismatches 152; Indels 144; Gaps 22;

Qy 19 KETAARAFERQHMDSPLDGTDDDDKAMADIGSIEGRGTMAISVTMDNLSGFENEYDVIY 78
Db 358 KENIKKIAKSAMEDVNSMDSDLNALISELYNIESR-----MDQIQRNLKLLKDLVE 409
Qy 79 LKPLAGVYRSUK-----KQIEKNIFTFN----- 101
Db 410 LKKEAQLENAKELDILEYKPIWVTRNKKHKNELFKNSYERKYEKEIYNOAEDELK 469
Qy 102 -LNLDILNSLRKKRYFLDVLESQMFKHIS-----SNEYIIE-----DSFKLL 146
Db 470 SMNISNL--DQYDKAKIIFDDKEQELIEAENIASLDAEKENSIIKYKAKRDDIIKQ 527
Qy 147 NSEQNTLLASYYIKESVENDIRFAQEGISYYEKVLAK-----YKDDLESIKKVIKE 199
Db 528 NSDILVNMETI-FADEVKEDSSKKIQTIN--EKELKKDENGINKENELKENKDIVKE 584
Qy 200 E-----KEKFPSPPTTPPSPAKTDEQKESK--FLPFLTNIETLYNNLVNKID-- 246
Db 585 ESLPKENVNKKQEDNKTVPADNKNKQYKKNKNNKSVKNGOSKNNTOKNPKOKKTL 644
Qy 247 -----DYVLINKAK-----INDCNVEKDE--AHVKITKL-SDLK----- 277
Db 645 EQLKKEKLLCSLKAKDELKISKNTISEDMAFCEKVKKIDEDIKILDKETISGLVGLKS 704
Qy 278 AIDDKIDLFK-----NPYDFAIKKLINDDTKMDLGLKLLSTGLVQNFPTIISK 327
Db 705 EIDDKINMKEKFFVSKMYNPIQDERSRYN--SLRENOIREITERAMAN-----MDK 757
Qy 328 LIEGKFOOM--LNISOHCQYKQCPENSGCFRHLDERECKCLLNYKQEGDKVCE 380
Db 758 AIVEEFKDISKLHTIENRIVEVQ-----KVLDLKQO-----MIDLEEGRKLKEE 801

RESULT 25
S51972
SPC72 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: FUN42 protein; protein YAL047c; STU2 protein interactant
C:Species: Saccharomyces cerevisiae
C:Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C:Accession: S51972
R:Bussey, H.; Kaback, D.B.; Zhong, W.; Vo, D.T.; Clark, M.W.; Fortin, N.; Hall, J.; O. submitted to the EMBL Data Library, August 1994
A:Description: The sequence of chromosome 1 of Saccharomyces cerevisiae.
A:Reference number: S51956
A:Accession: S51972
A:Molecule type: DNA
A:Residues: 1-622 <BUS>
A:Cross-references: EMBL:U12980; NID:g1326053; PID:g5955537; GSPDB:GN00001; MIPS:YAL04
C:Genetics:
A:Gene: SGD:SPC72; SPC72; FUN42; SPI6; MIPS:YAL047C
A:Cross-references: MIPS:YAL047c; SGD:S0000045
A:Map position: 1L

Query Match 6.4%; Score 146.5; DB 2; Length 622;
Best Local Similarity 18.7%; Pred. No. 2;
Matches 80; Conservative 75; Mismatches 125; Indels 147; Gaps 18;

Qy 18 MKETAARAFERQHMDSPLDGTDDDDKAMADIGSIEGRGTMAISVTMDNLSGFP----- 71
Db 228 LAQSSPAGSQLESRSPSKEENTDGGYQNDIHSNNHIDTENVMANSTSLPISAVESR 287
Qy 55 GTMAISVTMDNLSGFENEYDVIYKPLAGVYRSUKKQIEKNIFTFNLDILNSLRKK 114
Db 288 FEKTLQVLESDLMQFKHISNEYIIEFSKLLNSEQNTLKSYKIESVENDIRFAQEG 334
Qy 115 RYKFLDVLESDLMQFKHISNEYIIEFSKLLNSEQNTLKSYKIESVENDIRFAQEG 174
Db 335 QSHLLDSLELEENSSSVIEKQDHLISQLKEI--ESOSVLNNLEKLEKEDI--IKMKQN 389

[illegible]

Query Match	6.3%	Score 142.5;	DB 2;	Length 864;
Best Local Similarity	22.8%	Pred. No. 4.7;		
Matches	87;	Conservative 63;	Mismatches 150;	Indels 81; Gaps 16;
Qy	27	ERQHMDSPDGLGTDDDD-----KAMADIGSIEGRGTWAISVTMDNILSGF	70	
Db	364	ERLEKDAEUSLNDIDKVNLSLEQVETRTKKOLNRAQLAKVE--SLISEKNEIINNISQV	421	
Qy	71	ENEYDVATYLPACGYVR-SLKKQJKEKNITFNMLNLDILNSLRKKRYFYFDLVEESDLMOF	129	
Db	422	EGETCPVCGRPDLDEEHKQIIEAKSYILOLELNKNE-LEBELAKIITNELNKIE--REY	477	
Qy	130	KHISNSNEYITBD----SPKLSNSEQKN--TLKSYKYIKESVENDIKFAQEGISYYEKV--	182	
Db	478	RRLSNKASVDNVMRQLKMLNEETENLHSETESLUNDEEIKKINEEVKELKLYEYEFMR	537	
Qy	183	LAKY-KODLESIK-----KWIKEEK-----FPSPPTTTPSPAKTD	219	
Db	538	LSKYTKBELDKRRVKLDEMKKKEEIKEMKRGLESELKGDRALESKILDLENKRYKLD	597	
Qy	220	EOKK-----ESKFLPFLTNI---ETLYNNLVYNKIDDYILINLAKKIND	258	
Db	598	EMKKEKKGILEDYIROVKLLQEEVNLREEVNLIQFDENRYNELKTSYDANLSLKEKEN-	656	
Qy	259	CNVEKDEAHVKITKL-SDLKAIDDKIDLFKNPYPDFEATKLLINDDTKKMDLUGKLLISTGLV	317	

```

Query Match      6.3%; Score 143; DB 2; Length 540;
Best Local Similarity 22.5%; Pred. No. 2.6;
Matches 103; Conservative 53; Mismatches 147; Indels 154; Gaps 22;

Qy 9 SGLVPRSGMKETAAAKFERQHMDSPDLGTD-----DDKAMADIGSIEGRGTMAI 59
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 68 SNNVPFAEDLLET---YEHYH-DRPTLFLEDSWDGLKAKSNDLLSLGTSQTFRSNRE 122
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 60 SVTHMDNLS---GFENEYDVIIYKPLAGVYVRSLKKKQIEKNFTFFNLNDILNSRLKKR 115
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

C:Superfamily: hypothetical protein MJ1322						
Query Match		6.2%; Score 142; DB 2; Length 1005;				
Best Local Similarity		22.7%; Pred. No. 5.9;				
Matches 82; Conservative 68; Mismatches 117; Indels 94; Gaps						
Qy	63	MNTLSGFENEYD-VIYLKPLAGVYRSLLKKQIEKNITFTNLNNDILNSRLKK-----RK 116				
Db	346	LDTLTKIKDIERTVIKOLLBELKNLNNEIEK-----IKYKRICECKE 392				
Qy	117	YFLDVSLDMQFKHSSVEIIEDSPKLNSQK-----NTLLSKSYKVK- 162				
Db	393	YYEKYLE---LEEKAVEYNKLTLE--YITLQEKKSIEKNINDLETINKLLEETKNIDI 447				
Qy	163	ESVENDKFAQEGISYE-----KVLAHYKDDESIIKKVKEEKEKFPSSPTTPP 212				
Db	448	ESIENSLKEITEERKKVLLENLQEKELINLKVLGEINSEIKRLLKILDDELVEGKCPLCT 507				
Qy	213	PSPAKTDEOKESKFLPFNLTIETLYNNLVNKIDDYLINLKAINDCNVEK-----DEA 266				
Db	508	P----IDENKK---MELINQHKTQLNKKYTELEE---INKRIEIKDLKKKEIDKEE 557				
Qy	267	HVK-----ITKLSDLKAIDKIDLKPNDYFPAIKLIN-----DDTKMDMLGKLL 312				
Db	558	NLKTLLTYLEKQSIOEELELKLKNYKEQLD-EINKISNYVINGKPVDEILEDIKSQL- 615				
Qy	313	STGLVONFPWTISKL-----LEGFDQMLNISHOQVKKOCPENSGCFRHLDER 362				
Db	616	--NKFKNFYNOYSAVSYLNSVDDEGINRKIKEINIVS-GWNKCEKREELNKLRE-DER 671				
Qy	363	E 363				
Db	672	E 672				
RESULT 36						
Tl8355						
hypothetical protein P3 - Mycoplasma hyorhinis						
C:Species: Mycoplasma hyorhinis						
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 07-Dec-1999						
A:Accession: Tl8355						
R:Deng, G.; McIntosh, M.A.						
J. Bacteriol. 176, 5929-5937, 1994						
A>Title: An amplifiable DNA region from the Mycoplasma hyorhinis genome.						
A:Reference number: Z18888; MUID:95014025; PMID:7928953						
A:Accession: Tl8355						
A>Status: preliminary; translated from GB/EMBL/DDBJ						
A:Molecule type: DNA						
A:Residues: 1-1187 <DN>						
A:Cross-references: EMBL:L11447; NID:g150156; PID:g664956; PIDN:AAA62228.1						
C:Genetics:						
A:Genetic code: SGC3						
Query Match		6.2%; Score 142; DB 2; Length 1187;				
Best Local Similarity		20.7%; Pred. No. 7.1;				
Matches 90; Conservative 61; Mismatches 138; Indels 146; Gaps						
Qy	74	YDIVIYLKPLAGVY-----RSUKKQTE-----KNIFTNLNLDILNSRLKK 114				
Db	25	YRLRYAKKHVGIDNFHDIVVAINSLNAPARLWVEDNNFRVRFLREFPLNVNKLTTT--QR 82				
Qy	115	RKYFLDVLESOLMQFKHSSVEIIEDSPKLLNSEQNKTLKSKYKIKESVENDIKFAQE 174				
Db	83	LEFYNDVF-SDR-----KELSNSVELAEQQFELAEKNKRTLELAK----- 123				
Qy	175	GISYVERVLAKYKDDLESIKKVIKEE-----KPPSSPTTPPSPARTDBQKKE 224				
Db	124	-IDSLEKI-QELQEDTESIIISTTEEPQEBHQECETCRCLASSKPKVEATLEGEVVEQIV 181				
Qy	225	SKFLPFLTNITLYNNLVNK-----IDDYLI-----NLKAKINDCNVEKDE 265				
Db	182	SE-EPVVYLKEETVEKIVSNKECLSKILLDEVISTVCQGCAANSKWVCDLCESCNGQCHE 240				

Qy 90 KKQIEKNITFN-----LNNDILNSRLKKRYFLDVLESDLMQFKHISSEYI- 138
| | : : | : | : | : | : | : | :
Db 333 KSNINNNLYQLNSLLKEKKNDFSLSDEN--RYTKSFELVDLILSVLSAKSEFVL 389

Qy 139 ----IEDSFKILLNSEOKNTLLKSYYIKESVENDIKFAOEGISYYEKVLAKYKDOL- 191

Db 390 LKEHLDSKLFERS-----LSIKYIKERIVNIKY-----ISKDENLALLRDKIEPIF 439
Qy 192 ----SIRKVIKEEKEKFPSPPTPPSPAKTDEOKESKFLPFLTNIETLYNNLVNKIDD 247
Db 440 DONVNLKFFLEKN-----ALAKSLAKE-----ITNTERLIEEKTQLND 479
Qy 248 YLINLAKAKINDCNVEKDEAHVKIYK---LSDLKAIDDKIDLFKNPYDPEATKKLINDDTK 304
Db 480 VLGELE-----YIELSKFNLDIKLIDGNLEFL-----FESKNSL--DEEL 519
Qy 305 KDMGLKLLSTGLVONFPNTIISKLI-----EGKFODMLNISOHQCVKKQCPENS 353
Db 520 KDLYLKLLENLKSDDQLNLSNIGASKFSSESFEKEDFNSLVFLNDF-----KKT 571
Qy 354 GCFRHLDERECKLLANYKQGDCKVCNPNPTCNENGGCDADATCTEEDSGSS---RKK 410
Db 572 NYEYVIRKQTEYEFLLSNVSIKSEIENFN-VSNEM-----SDKFELEEDLATRELLRKE 625
Qy 411 ITCECTRPDSYPLFDGIF 428
Db 626 I-----DAIKLGDYVF 636

RESULT 40

C89910
hypothetical protein SAI181 [Imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: C89910
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, M.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: C89910
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1009 <KUR>
A:Cross-references: GB:BA000018; PID:gl3701144; PIDN:BA42439.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SAI181

Query Match 6.28; Score 140; DB 2; Length 1009;
Best Local Similarity 22.98; Pred. No. 7.5;
Matches 97; Conservative 51; Mismatches 131; Indels 144; Gaps 21;

Qy 19 KETAAKFERQHMD--PDLCTDDDDKAMADIGSIEGRGTWASVTMDNI-----L 67
Db 193 KEKAIENRYQOIDLWQIESEFDDK----IKGLLELATQOIKLIENIPLLQARSKEI 248
Qy 68 SGFEN-----EYDVYILKPLAGVYRSKQIEKNIFTNLDNLNLSRLKRRYFL 119
Db 249 LAFVNESKETAKEVEITE-----KKTLENNILKDNIN-----QLNKNKI-- 288
Qy 120 DVLESDLMPKHSSNEYIIEDSFKLLNSEQNTLLSKYIKESVENDIKFAQEGISYY 179
Db 289 -----DFVLKEQPEIDEIEAKLKL---QDITNLLNYIENREKIETKIANSKKDISKT 340
Qy 180 EKLAKY---KDDLESIKKVIKEKEKFPSPPTTPSPAKTDEOKESKFLPFLTNIET 236
Db 341 NKLINLDCDRNDKCKMLLENGDLIES-----KTS-----FIDKTRV 380
Qy 237 LYNNLVNKIDDYLYNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDPEAIK 296
Db 381 LFND-INKYQOSYLNIECLITEGQLGDELN---NLIKLEKVEDSIG--NNESDYEKII 434
Qy 297 KLIN-----DDTKKDMGLKLLSTGLVQNFN-----TIISKLIEGKFQDML 337
Db 435 ELNNAITNINNEINIENKAKAEKLDKLL--GSKQLENOINEETTMMKNLEIK----- 487

Qy 338 NISOHQCVKKQCPENSGCFRHLDERECKCLLNYKQE-----GDKCVENPNPTC 386
Db 488 -----LDHYDKSKLDLNDKESFISEIKSAVKIGDQD-----PIC 521
Qy 387 -NE 388
Db 522 GNE 524

Search completed: May 12, 2003, 10:20:30
Job time : 70 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 10:09:22 ; Search time 25 Seconds
(without alignments)
715.052 Million cell updates/sec

Title: US-10-057-531A-2
Perfect score: 2275
Sequence: 1 MHHHHSSGLVPRGSMKE.....TCECTKPDSPFLDGFICSS 431

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1946.5	85.6	1701	1 MSPI_PLAFF	P13819 plasmodium
2	1946.5	85.6	1701	1 MSPI_PLAFF	P08569 plasmodium
3	1935.5	85.1	1726	1 MSPI_PLAF3	P04934 plasmodium
4	1935	85.1	1682	1 MSPI_PLAF3	P15598 plasmodium
5	1932.5	84.9	1726	1 MSPI_PLAF3	P50495 plasmodium
6	1175	51.6	1630	1 MSPI_PLAFK	P04932 plasmodium
7	1175	51.6	1639	1 MSPI_PLAFW	P04933 plasmodium
8	556.5	24.5	1772	1 MSPI_PLAYO	P13828 plasmodium
9	160.5	7.1	1169	1 SMC_METJA	Q59037 methanococc
10	146.5	6.4	622	1 YAE7_YEAST	P39723 saccharomyc
11	145	6.4	1130	1 YL17_CAEEL	Q11102 caenorhabdi
12	143	6.3	540	1 YK26_YEAST	P36112 saccharomyc
13	142.5	6.3	864	1 RASO_SULSO	Q97900 sulfolobus
14	142	6.2	1005	1 RASO_METJA	Q58718 methanococc
15	142	6.2	1679	1 Y109_YEAST	P40457 saccharomyc
16	139.5	6.1	879	1 RASO_SULTO	Q96495 sulfolobus
17	138.5	6.1	2116	1 MY52_DICDI	P08799 dictyostell
18	136	6.0	1163	1 SBCC_CLOAB	Q97181 clostridium
19	135.5	6.0	2869	1 RBP1_PLAVB	Q00798 plasmodium
20	135	5.9	880	1 RASO_PYRAB	Q90498 pyrococcus
21	134	5.9	1790	1 USOL_YEAST	P25386 saccharomyc
22	134	5.9	1928	1 MY51_YEAST	P08564 saccharomyc
23	133.5	5.9	1170	1 SMC2_YEAST	P38989 saccharomyc
24	133	5.8	978	1 RASO_AQUAE	Q67124 aquifex ae
25	133	5.8	2104	1 MY53_SCHPO	Q14157 schizosacch
26	132.5	5.8	886	1 RASO_SULAC	Q33600 sulfolobus
27	132	5.8	882	1 RASO_PYRFU	P58301 pyrococcus
28	132	5.8	3329	1 BRC2_MOUSE	P97929 mus musculu
29	131.5	5.8	707	1 YJ9C_YEAST	P47166 saccharomyc
30	131.5	5.8	1225	1 SMC1_YEAST	P32908 saccharomyc
31	131	5.8	1251	1 RBP2_PLAVB	Q00799 plasmodium
32	130.5	5.7	895	1 RASO_THEVO	P58302 thermoplasm
33	130	5.7	886	1 RASO_ARCFU	O29230 archaeoglob

RESULT 1				
MSPI_PLAFF	STANDARD;	PRT;	1701 AA.	
ID	MSPI_PLAFF			
AC	P13819;			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Merozoite surface protein 1 precursor (Merozoite surface antigens)			
DE	(PMMSA).			
GN	MSP-1.			
OS	Plasmodium falciparum (isolate FC27 / Papua New Guinea).			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_Taxid=5837;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88142999; PubMed=2449612;			
RA	Peterson M.G. Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,			
RA	Brown G.V., Anders R.F., Kemp D.J.;			
RT	"Variation in the precursor to the major merozoite surface antigens			
RT	of Plasmodium falciparum."			
RL	Mol. Biochem. Parasitol. 27:291-302(1988).			
CC	-1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor			
CC	(Potential)			
CC	-1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42			
CC	kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF			
CC	MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL: M19143; AAA29653.1; -			
DR	PIR: A54498; A54498. EGF-like.			
DR	InterPro: IPR000561; EGF-like.			
DR	Pfam: PF00008; EGF; 1.			
KW	Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;			
KW	Transmembrane; GPI-anchor.			
FT	SIGNAL 1 19	POTENTIAL		
FT	CHAIN 20 1701	MEROZOITE SURFACE PROTEIN 1.		
FT	CARBOHYD 110 110	N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 239 239	N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 470 470	N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 536 536	N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 607 607	N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 802 802	N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 899 899	N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 919 919	N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 965 965	N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 991 991	N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 1089 1089	N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 1196 1196	N-LINKED (GLCNAC. .) (POTENTIAL).		
FT	CARBOHYD 1588 1588	N-LINKED (GLCNAC. .) (POTENTIAL).		

P47037 saccharomyc
P47460 mycoplasma
Q9wva1 thermotoga
P32380 saccharomyc
P27895 saccharomyc
P46468 plasmodium
O58687 pyrococcus
Q13201 homo sapien
Q95293 dictyostelli
P06101 saccharomyc
P17119 saccharomyc
P43573 saccharomyc

ALIGNMENTS

```
SQ SEQUENCE 1701 AA; 193719 MW; 3920B75E73D38552 CRC64;
Query Match 85.6%; Score 1946.5; DB 1; Length 1701;
Best Local Similarity 94.9%; Pred. No. 9.8e-93;
Matches 374; Conservative 7; Mismatches 8; Indels 5; Gaps 1;

QY 38 TDDDDKAMADIGSTEGRTMAISVTMDNILSGFENEYDVIYKPLAGYVRSLLKQKQIEKNI 1347
Db 1293 SEDNDEYLDQVVTGE-----AISVTMDNILSGFENEYDVIYKPLAGYVRSLLKQKQIEKNI 1347

QY 98 FTFNLNLDILNLSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSFKLNSEQKNTLLKS 157
Db 1348 ITFNLNLDILNLSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSFKLNSEQKNTLLKS 1407

QY 158 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKFPSPPTPPSPAK 217
Db 1408 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKFPSPPTPPSPAK 1467

QY 218 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYILNKLAKINDCNVEKDEAHVKITKLSDLK 277
Db 1468 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYILNKLAKINDCNVEKDEAHVKITKLSDLK 1527

QY 278 AIDDKIDLFKNPYDPEATKLLINDTKMDLGLKLLSTGLVONFNTIISKLECKFODML 337
Db 1528 AIDDKIDLFKNPYDPEATKLLINDTKMDLGLKLLSTGLVONFNTIISKLECKFODML 1587

QY 338 NISOHQCVKKQCPENSGCFRHLDERECKCLLYNKQEGDKCVENPNPTCNENNGCCDADA 397
Db 1588 NISOHQCVKKQCPENSGCFRHLDERECKCLLYNKQEGDKCVENPNPTCNENNGCCDADA 1647

QY 398 TCTEEDSGSSRRKKTCTCKTPDPSYPLFDGIFCSS 431
Db 1648 TCTEEDSGSSRRKKTCTCKTPDPSYPLFDGIFCSS 1681

RESULT 2
MSPL_PLAFM STANDARD; PRT; 1701 AA.
ID MSPL_PLAFM AC P08569;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMWSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=70153;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88011243; PubMed=3079521;
RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT "Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum";
RT Plasmodium falciparum;
RL J. Mol. Biol. 195:273-287(1987).
RN [2]
RP REVISIONS TO 1403; 1569 AND 1629.
RA Tanabe K.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-115 FROM N.A.
RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RT Stunnenberg H., Buyard H.;
RT "Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
CC CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (potential).
CC CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC CC -----
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X05624; CAA29112.1; -
DR PIR: A26868; A26868.
DR BIR: B25120; B25120.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96EA98 CRC64;
Query Match 85.6%; Score 1946.5; DB 1; Length 1701;
Best Local Similarity 94.9%; Pred. No. 9.8e-93;
Matches 374; Conservative 7; Mismatches 8; Indels 5; Gaps 1;

QY 38 TDDDDKAMADIGSTEGRTMAISVTMDNILSGFENEYDVIYKPLAGYVRSLLKQKQIEKNI 97
Db 1293 SEDNDEYLDQVVTGE-----AISVTMDNILSGFENEYDVIYKPLAGYVRSLLKQKQIEKNI 1347

QY 98 FTFNLNLDILNLSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSFKLNSEQKNTLLKS 157
Db 1348 ITFNLNLDILNLSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSFKLNSEQKNTLLKS 1407

QY 158 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKFPSPPTPPSPAK 217
Db 1408 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKFPSPPTPPSPAK 1467

QY 218 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYILNKLAKINDCNVEKDEAHVKITKLSDLK 277
Db 1468 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYILNKLAKINDCNVEKDEAHVKITKLSDLK 1527

QY 278 AIDDKIDLFKNPYDPEATKLLINDTKMDLGLKLLSTGLVONFNTIISKLECKFODML 337
Db 1528 AIDDKIDLFKNPYDPEATKLLINDTKMDLGLKLLSTGLVONFNTIISKLECKFODML 1587

QY 338 NISOHQCVKKQCPENSGCFRHLDERECKCLLYNKQEGDKCVENPNPTCNENNGCCDADA 397
Db 1588 NISOHQCVKKQCPENSGCFRHLDERECKCLLYNKQEGDKCVENPNPTCNENNGCCDADA 1647

QY 398 TCTEEDSGSSRRKKTCTCKTPDPSYPLFDGIFCSS 431
Db 1648 TCTEEDSGSSRRKKTCTCKTPDPSYPLFDGIFCSS 1681

RESULT 3
MSPL_PLAFM STANDARD; PRT; 1726 AA.
ID MSPL_PLAFM AC P04934;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
```


FT CARBOHYD 881 881 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 947 947 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1178 1178 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1569 1569 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1682 AA; 192462 MW; C82A1E159948CAD6 CRC64;

Query Match 85.1%; Score 1935; DB 1; Length 1682;
 Best Local Similarity 94.4%; Pred. No. 3.8e-92;
 Matches 372; Conservative 9; Mismatches 7; Indels 6; Gaps 2;

QY 38 TDDDDKAMADIGSTEGRTMAISVTMONILSGFENEVDVYVYKPLAGVYRSLKQKQIEKNI 97
 DB 1275 SEDNEVDLQVVTGE-----AISVTMONILSGFENEVDVYVYKPLAGVYRSLKQKQIEKNI 1329
 QY 98 FTFNLNLDILNLSKRRKRYFLDVLSDLMQFKHISNEYIIESDFKLLNSEQKNTLLKS 157
 DB 1330 FTFNLNLDILNLSKRRKRYFLDVLSDLMQFKHISNEYIIESDFKLLNSEQKNTLLKS 1389
 QY 158 YKIKESVENDIKFAQEGISYERVLAKYKDDLESIKKVIKEKEKFPSPPTTTPSPAK 217
 DB 1390 YKIKESVENDIKFAQEGISYERVLAKYKDDLESIKKVIKEKEKFPSPPTTTPSPAK 1448
 QY 218 TDEQKESKFLPFTNLTETLYNNLVNKIDDYLYNLKAKINDCNVEKDEAHVKITKLSDLK 277
 DB 1449 TDEQKESKFLPFTNLTETLYNNLVNKIDDYLYNLKAKINDCNVEKDEAHVKITKLSDLK 1508
 QY 278 AIDDKIDLFKNPYDFEAIKLLNDTKKMLGKLLSTGLVQNFNTIISKLECKFQDML 337
 DB 1509 AIDDKIDLFKNPYDFEAIKLLNDTKKMLGKLLSTGLVQNFNTIISKLECKFQDML 1568
 QY 338 NISQHCVKKQCPNSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCCDADA 397
 DB 1569 NISQHCVKKQCPNSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCCDADA 1628
 QY 398 TCTEEDSGSSRKKITCCTKTPDPSYPLFDGIFCSS 431
 DB 1629 KCTEEDSGSGKKITCCTKTPDPSYPLFDGIFCSS 1662

RESULT 5
 MSPL_PLAFK ID MSPL_PLAFK STANDARD; PRT; 1726 AA.
 AC P0495;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Merzoite surface protein 1 precursor (Merzoite surface antigens) (PMMSA) (GP195).
 GN MSP-1.
 OS Plasmodium falciparum (isolate Palo Alto / Uganda).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=57270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89005525; PubMed=3049134;
 RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
 RA Siddiqui W.A.;
 RT "Plasmodium falciparum: gene structure and hydrophobic profile of the
 RT major merzoite surface antigen (gpi95) of the Uganda-Palo Alto
 RT isolate.";
 RL Exp. Parasitol. 67:1-11(1988).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (potential).
 CC -1- PTM: MERZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MERZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M37213; AAA29611.1; --
 CC InterPro: IPR000561; EGF-like.
 CC Pfam: PF00008; EGF; 1.
 CC Malaria; Merzoite; Polyprotein; Repeat; Signal; Glycoprotein;
 CC Transmembrane; GPI-anchor.
 CC SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1726 MERZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1726 AA; 196174 MW; 5B59CBEFA2F9A026 CRC64;

Query Match 84.9%; Score 1932.5; DB 1; Length 1726;
 Best Local Similarity 94.2%; Pred. No. 5.2e-92;
 Matches 371; Conservative 9; Mismatches 9; Indels 5; Gaps 1;

QY 38 TDDDDKAMADIGSTEGRTMAISVTMONILSGFENEVDVYVYKPLAGVYRSLKQKQIEKNI 97
 DB 1318 SEDNEVDLQVVTGE-----AISVTMONILSGFENEVDVYVYKPLAGVYRSLKQKQIEKNI 1372
 QY 98 FTFNLNLDILNLSKRRKRYFLDVLSDLMQFKHISNEYIIESDFKLLNSEQKNTLLKS 157
 DB 1373 FTFNLNLDILNLSKRRKRYFLDVLSDLMQFKHISNEYIIESDFKLLNSEQKNTLLKS 1432
 QY 158 YKIKESVENDIKFAQEGISYERVLAKYKDDLESIKKVIKEKEKFPSPPTTTPSPAK 217
 DB 1433 YKIKESVENDIKFAQEGISYERVLAKYKDDLESIKKVIKEKEKFPSPPTTTPSPAK 1492
 QY 218 TDEQKESKFLPFTNLTETLYNNLVNKIDDYLYNLKAKINDCNVEKDEAHVKITKLSDLK 277
 DB 1493 TDEQKESKFLPFTNLTETLYNNLVNKIDDYLYNLKAKINDCNVEKDEAHVKITKLSDLK 1552
 QY 278 AIDDKIDLFKNPYDFEAIKLLNDTKKMLGKLLSTGLVQNFNTIISKLECKFQDML 337
 DB 1553 AIDDKIDLFKNPYDFEAIKLLNDTKKMLGKLLSTGLVQNFNTIISKLECKFQDML 1612
 QY 338 NISQHCVKKQCPNSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCCDADA 397
 DB 1613 NISQHCVKKQCPNSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCCDADA 1672
 QY 398 TCTEEDSGSSRKKITCCTKTPDPSYPLFDGIFCSS 431
 DB 1673 KCTEEDSGSGKKITCCTKTPDPSYPLFDGIFCSS 1706

RESULT 6
 MSPL_PLAFK ID MSPL_PLAFK STANDARD; PRT; 1630 AA.
 AC P04932;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Merzoite surface protein 1 precursor (Merzoite surface antigens) (PMMSA) (P190).
 GN MSP-1.
 OS Plasmodium falciparum (isolate K1 / Thailand).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5839;
 RN [1]

SEQUENCE FROM N.A.
 RX MEDLINE=86136024; PubMed=3004972;
 RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
 RA Stunnenberg H., Bujard H.;
 RT "Polymorphism of the precursor for the major surface antigens of
 RT Plasmodium falciparum merozoites: studies at the genetic level.";
 RL EMBO J. 4:3823-3829(1985).
 RN [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RA Pan W., Tolle R., Bujard H.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (potential).
 CC PTN: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X03371; CAA27070.1; -;
 DR PIR; A25120; SAZ0K1.
 DR InterPro: IPR000561; EGF-like.
 DR Pfam: PF00008; EGF; 1
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19
 FT CHAIN 20 1630
 FT DOMAIN 67 84
 FT TRANSMEM 1614 1630
 FT CARBOHYD 97 97
 FT CARBOHYD 259 259
 FT CARBOHYD 755 755
 FT CARBOHYD 759 759
 FT CARBOHYD 774 774
 FT CARBOHYD 835 835
 FT CARBOHYD 911 911
 FT CARBOHYD 955 955
 FT CARBOHYD 1049 1049
 FT CARBOHYD 1156 1156
 FT CARBOHYD 1165 1165
 FT CARBOHYD 1436 1436
 FT CARBOHYD 1517 1517
 FT SEQUENCE 1630 AA; 187289 MW; ADBDEC3CE0A46322 CRC64;
 Query Match 51.6%; Score 1175; DB 1; Length 1630;
 Best Local Similarity 56.6%; Pred. No. 3.6e-53;
 Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;
 QY 34 POLGTDHDDKADMGSTEGRTMAISVTMDNLSGFENYDVLYKPLAGVYRSLKQI 93
 DB 1235 PLFGSEED--YDDLGVQVGTGAVTPSV-IDNLSKIENEYEVLYKPLAGVYRSLKQI 1291
 QY 94 EKNIETNLNLDILNRLKRYKFLDVLSDLMQFKHSSNEYIEDSFKLLNSEQKNT 153
 DB 1292 ENNVTFNVNVDILNRFNKNENKFNVLSDLPYKDTSSNVYVVDYKFLNKEKRD 1351
 QY 154 LKSKYIKESVENDIKFAQEGISYERKVLAKYKDDLESIKKVIKEKEKFPSPPTTP 213
 DB 1352 FLSSYNYIKSDITDINFANDVLYGYKILSEKYKSLDLSIKYI----- 1395
 QY 214 SPAKTDQKESKFLPFTTETLTNNLNNLKDIDYLLNKKINDCNVEKDEAHVKYTKL 273
 DB 1396 ----NDKQGENEKYLPFLNNIETLYKTVDKIDLEVIHLEAKVLNVTYKSNVVKIKEL 1451
 QY 274 SDLKAIDDKIDLFPKNPYDFEAKKLINDTKKMLGKLLSTGLV-QNFPNTIISKLECK 332
 DB 1452 NYLKTIQDKLADFKNNNFVGIADLSTQDYNHNNLLTKFLSTGMVFENLAKTVLSNLLDGN 1511

QY 333 FODMLNISQCVKQKOCPEKSCFRLHDERECKCLLNKQEGDKCVENPNTCNENGG 392
 DB 1512 LOGMLNISQCVKQKOCPEKSCFRLHDERECKCLLNKQEGDKCVENPNTCNENGG 1571
 QY 393 CDADATCTEEDSGSSRRKITCTCKPDSYPLFDGIFCSS 431
 DB 1572 CDADAKCTEEDSGSNGKKTCTCKPDSYPLFDGIFCSS 1610
 RESULT 7
 MSPI_PLAFW
 ID MSPI_PLAFW STANDARD; PRT; 1639 AA.
 AC P04933;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMMSA) (P195).
 GN MSP-1.
 OS Plasmodium falciparum (isolate Wellcome).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OC NCBI_TaxID=5848;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86014355; PubMed=2995820;
 RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
 RA Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
 RA Freeman R.N.;
 RT "Primary structure of the precursor to the three major surface
 RT antigens of Plasmodium falciparum merozoites.";
 RL Nature 317:270-273(1985).
 RN [2]
 RP REVISIONS.
 RA Holder A.A.;
 RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (potential).
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X02919; CAA26676.1; -;
 DR PIR; A24594; A24594.
 DR InterPro: IPR000561; EGF-like.
 DR Pfam: PF00008; EGF; 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19
 FT CHAIN 20 1639
 FT CARBOHYD 116 116
 FT CARBOHYD 268 268
 FT CARBOHYD 764 764
 FT CARBOHYD 768 768
 FT CARBOHYD 783 783
 FT CARBOHYD 844 844
 FT CARBOHYD 920 920
 FT CARBOHYD 964 964
 FT CARBOHYD 1058 1058
 FT CARBOHYD 1165 1165
 FT CARBOHYD 1174 1174
 FT CARBOHYD 1445 1445
 FT CARBOHYD 1526 1526
 FT SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;
 SQ

```
Query Match 51.6%; Score 1175; DB 1; Length 1639;
Best Local Similarity 56.6%; Pred. No. 3.6e-53;
Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;

QY 34 PDLCTDDDDKAMADIGSIEGRGTWAIWTDNIIISGENEYDVIIYKPLAGVYSLKQOI 93
DB 1244 PIFGESEED--YDLDGVVGTGEAVTPSV-IDNIIISKIENEYEVLYKPLAGVYSLKQOL 1300
QY 94 EKNIFTNENLNDLSRLKRRKFLDVLSDLMQFHHISSNEYIIEDSFKLLNSEOKNT 153
DB 1301 ENNVWTVNVVNDILNRFKRNKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNK 1360
QY 154 LLKSYKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVIKKEKFPSPPTTPP 213
DB 1361 FLSSYNIKIDSIDYINFANDVILGYKILSEKYSKLDLSIKKI- 1404
QY 214 SPATDQKESKFLPFLTNTIETLYNNLVKIDYLLNLRKAKINDCNVEXDEAHVKITKL 273
DB 1405 ----NDQGENEKYLPENNIETLYKTVNOKIDLVFVHLEAKVLYNYTESNVEVKIKEL 1460
QY 274 SLDKAIKDDKIDLFKNPYDFAIKKLLINDTKMDGLKLLSTGLV-QNFPNTIISKLIIEGK 332
DB 1461 NYLTIQDKLADFPKNNFNKFNKFNKFNKFNKFNKFNKFNKFNKFNKFNK 1520
QY 333 FQDMLNISQHCQVKKQPCNSGCRPHLDREBECKLLNYKQEGDKCVENPNTCNENNGG 392
DB 1521 LQGMNLISQHCQVKKQPCNSGCRPHLDREBECKLLNYKQEGDKCVENPNTCNENNGG 1580
QY 393 CDADATCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431
DB 1581 CDADAKTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 1619

RESULT 8
MSPL_PLAYO STANDARD; PRT; 1772 AA.
AC P13828;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMA) (230 kDa).
MSF-1.
OS Plasmodium berghei yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5862;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90205979; PubMed=2320061;
RA Lewis A.P.;
RT "Sequence analysis upstream of the gene encoding the precursor to the
RL major merozoite surface antigens of Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 39:285-288(1990).
RN [2]
RP SEQUENCE OF 1093-1772 FROM N.A..
RC STRAIN=17XL;
RX MEDLINE=88124889; PubMed=2448778;
RA Burns J.M. Jr., Daly T.M., Vaidya A.B., Long C.A.;
RT "The 3' portion of the gene for a Plasmodium yoelii merozoite surface
RT antigen encodes the epitope recognized by a protective monoclonal
RT antibody.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:602-606(1988).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```



```

OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RA MEDLINE=96337999; PubMed=868087;
RA Bult C.-J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RL jannaschii."
RN Science 273:1058-1073(1996).
RP [2]
RP REVISIONS.
RA Bult C.-J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RC !- FUNCTION: PLAYS AN IMPORTANT ROLE IN CHROMOSOME STRUCTURE AND
CC PARTITIONING. ESSENTIAL FOR CHROMOSOME PARTITION (BY SIMILARITY).
CC !- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67604; AAB99663.1; -
CC TIGR; M31643; -
CC InterPro; IPR003439; ABC_transportr.
CC InterPro; IPR005289; GTP-binding_dom.
CC InterPro; IPR003405; SMC_C.
CC InterPro; IPR003395; SMC_N.
CC Pfam; PF02463; SMC_N; 1.
CC Pfam; PF02483; SMC_C; 1.
CC TIGRFAMs; TIGR00650; MG442; 1.
KW Hypothetical protein; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 31 38 ATP (POTENTIAL).
FT DOMAIN 160 521 COILED COIL (POTENTIAL).
FT DOMAIN 673 1032 COILED COIL (POTENTIAL).
FT SEQUENCE 1169 AA; 136634 MW; B63CE34E4C03F36 CRC64;
SQ
Query Match 7.1%; Score 160.5; DB 1; Length 1169;
Best Local Similarity 24.2%; Pred. No. 0.28;
Matches 86; Conservative 59; Mismatches 110; Indels 101; Gaps 17;
QY 94 EKNITFNLNLDLSRLK---KRYFLDVLSDLMQFKHISSEYIIDSFKLL-----146
DB 85 ENNAFNVNADKVGILR-RIKSSGETDYLLWKNDEKRRKKMKTHEII--DLFRRLGLLG 141
QY 147 -NSEQNTLLK-----SYKYIKESVENDIKFAQEGISYVEKVLAK 185
DB 142 DNVISGDLKIINISPIERRKIIDEISGIAEFDEKKKAEELKKARELIEMIDIRISE 201
QY 186 YKODLESIKKVIKEKEKFPSPPTTPSPAKTDEQKESKFLPFL-----231
DB 202 VENNKLKKE-KEDAKEY-----IKLNEELKAAKAYALILKKVSYLVNLENIQ 249
QY 232 ---TNTETLYNNLVNK---IDDYLINKAKINDC-----NVEKDEAHVKITKL-----273

```

```

Db 250 NDIKNEELKNEFLSKVREIDVEIENLKLRLNNIINELNEKNGNEVLELHKSIKELEVEI 309
QY 274 -SDLKALDDKI-DLEKNPYDFAIKKLINDTTKMDLGLKLLSTGLVQNPNTIIS---K 327
DB 310 ENDKKVLDSINELKKEVEIENKKEIKETQKK-----IENRDSIIEKEQOIK 359
QY 328 LIEGKQDMLNISOHQVKKQCPENSGCFRHLDEREECKLLNKKQBGKCVENPN 383
DB 360 EIEEKIKN-LNVEKER-LKEAIAESESIKHLKESE-----METADBIANKQN 405
RESULT 10
YAE7_YEAST STANDARD; PRT; 622 AA.
ID YAE7_YEAST AC P39723;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 72.1 kDa protein in ACS1-GCV3 intergenic region.
GN YAL047C OR FUM42.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
RT cerevisiae."
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U12980; AAC04984.1; -
CC DR SGD; S0000045; SPC72.
CC KW Hypothetical protein.
CC SEQUENCE 622 AA; 72104 MW; 8A3248B8BE434D31 CRC64;
Query Match 6.4%; Score 146.5; DB 1; Length 622;
Best Local Similarity 18.7%; Pred. No. 0.7;
Matches 80; Conservative 75; Mismatches 125; Indels 147; Gaps 18;
QY 18 MKETAAAKFERQHMDSF---DLGTD-----DDDKAMAD-----IGSIEGR 54
DB 228 LAOSSPAGSOLSRDSSPSKEENTDGGYQYQNDIHSNNHIDTENVMANSTSLPSAVESR 287
QY 55 GTMAISVNTNLISGFEYEDVYLYKPLAGVYRSKQIEKNIETFNLDINLSRLKK 114
DB 288 FEXTLDQQLVIVIELLHKEVDQF-----INSIRLKEF-----SOKLEIATSKLNE 334
QY 115 RYFLDVLSDLMQFKHISSEYIIDSFKLLNSEQNTLLKSYKYIKESVENDIKFAQE 174
DB 335 QSHLLDSELEENSSSVSSVQDHLISQLKEKI--ESQSVLNNLEKLEKEDI---IKMKQN 389
QY 175 GISYVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAKTDEQKESKFLPFLTNI 234
DB 390 -----EKLTK---ELETQTKINKLKE-----408
QY 235 ETLYNNLVNKIDDDYLINKAKINDCNVEK-DEAHV---KITKL-----SDLKAIIDDK 282
DB 409 -----NWDVSINDLEKQINDLQIDKSEEFHVIONQLDKLDLENYOLKNQNTLDNQ 460
QY 283 IDLFKNPYDFAIK-----KLINDDTKKMDLGLKLLSTGLVQNF-----PNTI-----324

```



```
RESULT 13
RA50_SULSO          STANDARD;          PRT;      864 AA.
AC Q97WH0:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR S502249.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1].
SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aweyze M.J., Chan-Welher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RA Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
RL CC
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE006829; AAK42417.1;
CC InterPro; IPR003439; ABC transportr.
CC InterPro; IPR003405; SMC_C.
CC InterPro; IPR003395; SMC_N.
CC Pfam; PF02463; SMC_N; 1.
CC Pfam; PF02483; SMC_C; 1.
CC DNA repair; Hydrolyase; ATP-binding; Coiled coil; Complete proteome.
CC NP_BIND 30 37 ATP (BY SIMILARITY).
CC DOMAIN 164 701 COILED COIL (POTENTIAL).
CC SEQUENCE 864 AA; 101601 MW; 657076AE9B709FC CRC64;
Query Match
Best Local Similarity 6.3%; Score 142.5; DB 1; Length 864;
Matches 87; Conservative 63; Mismatches 150; Indels 81; Gaps 16;
QY 27 ERQHMDSPDLGTDDDD-----KAMADIGSIEGRTMTAIVTMDNLISGF 70
DB 364 ERLEKDAESELNDIDKVNLSLEQKVEETRRKKNLNLAQLAKVE--SLISEKNEIINNISQV 421
QY 71 ENEYDIVYLKPLAGVYR-SLKQIEKNIFTNLDNLNLDIINSLRKLKRYFLDVLESMLQOF 129
DB 422 EGCTPCVGRPLDEEHKQIKIEAKSYILQLELNKNE-LEELKTKITNELNKIE---REY 477
QY 130 KHISNEYIIE--SFKLNLSEQKN--TLKSYKYIKESVENDIKFAQEGISYVEKV-- 182
DB 478 RRLSNKASYDVMNVRQLKNEIENLHSEIESLKNIDEIKNEEVKELKLYEEFMR 537
QY 183 LAKY-KDDLESIK-----KVYKEEK-----FPSPPTTPPSAKTD 219
DB 538 LSKYTKKEELDKRVKLDENMKKKEETEKEMRGLESELKLDLRKALESKILDLNKRKVLKD 597
```

```
QY 220 EOKK-----ESKFLPFLTNI-----ETLYNNLVNKIDDYLINLRAKIND 258
DB 598 EMKKKGILEDIYRQVKLLQEEVKNLREVNIIQFDENRYNELKTSLDAYNLSKEKEN- 556
QY 259 CNVEKDEAHVKTTL-SDLKAIDDKIDLPKNPYDFEAIKKLINDDPKQMLKGLSLTGLV 317
DB 657 ---RKSRIEGELESLKEDTIEISNRIANYE--LQLKDREKIINAINKLEKIRSAICERKL 711
QY 318 QNFPNPTIISKLIIEGKFDQMLN 338
DB 712 QSYIITTKQLIENNLNDIIS 732
```

```
RESULT 14
RA50_METJA
ID RA50_METJA          STANDARD;          PRT;     1005 AA.
AC Q58718;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR MJ1322.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1].
SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii.";
RT Science 273:1058-1073(1996).
RL CC
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67572; AAB99331.1;
CC TIGR; MJ1322;
CC InterPro; IPR003439; ABC transportr.
CC InterPro; IPR003405; SMC_C.
CC InterPro; IPR003395; SMC_N.
CC Pfam; PF02463; SMC_N; 1.
CC Pfam; PF02483; SMC_C; 1.
CC ProDom; PD000006; ABC_transportr; 1.
CC DNA repair; Hydrolyase; ATP-binding; Coiled coil; Complete proteome.
CC NP_BIND 32 39 ATP (BY SIMILARITY).
CC DOMAIN 158 849 COILED COIL (POTENTIAL).
CC SEQUENCE 1005 AA; 119387 MW; 9BBB48173E78F3 CRC64;
Query Match
6.2%; Score 142; DB 1; Length 1005;
```



```
DR EMBL; AP000988; BAB67212.1; ALT_INIT.
DR InterPro; IPR003439; ABC transportr.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF00005; ABC_tran; 1.
DR Pfam; PF02483; SMC_C; 1.
DR Pfam; PF02463; SMC_N; 1.
DR ProDom; PD000006; ABC_transportr; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil;
FT NP_BIND 30 37 ATP (BY SIMILARITY).
FT DOMAIN 172 731 COILED COIL (POTENTIAL).
SQ SEQUENCE 879 AA; 103487 MW; E9103B58914F20D7 CRC64;

Query Match 6.1%; Score 139.5; DB 1; Length 879;
Best Local Similarity 24.8%; Pred. No. 2.4; Mismatches 119; Indels 63; Gaps 14;
Matches 78; Conservative 55;

QY 88 SLKKQIEKNIFTNLDNLNLSRLKRYFLDVLKSLQKHSNVEYIIEIDSFKLNL 147
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 250 SLNLKDDISLREEVKDNRLREKEKLEKIDLEKD;---KLIEKEKIIIEAQNKIKL 305

QY 148 SEQKNTLLSKYKIVKESVENDKFAQEGISYKVEKVLAKYKDDLESKKVIKEEKE-KFPS 206
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 306 AQEKEKSLTKINLTDLSEKLRKRELEEDYKKYI-EIKGELEL;-----BEKERKFN 359

QY 207 SPPTTPPSPAKTDQEKESKELPELTNIETLYNNL;-----VNKIDYLLNLKAKINDCNV 261
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 360 -----LSDRLSKLIK--LSEIESKISNRKISINIEELDEKLOKLNEDLNKKQ 406

QY 262 EKDEAHVKITKLSDKAIDKIDLFKNPYDFFAIKKLINDDTKK-----DMLGKLLSTGL 316
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 407 EREKL;---ASQLGEIKG-----RIEELNKLNLGNQVKNVCPVCGRELSDDH 451

QY 317 VQNFNPNTIISKL;-----IEGKFQDMLN;---ISQ-HQCVKKQCPNSCFPHLDEREC 365
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 452 KRKIQNEIEIKLELDELNKLKFKLEINKINGLUISLNQIINKSKKEKDIAIRNLADYNN- 510

QY 366 KCLLNYKQEGDKCQE 380
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 511 --LLTQQQLRKEIE 523

RESULT 17
ID MYS2.DICD1
AC P08799;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin II heavy chain, non muscle.
GN MHCA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxId=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87092266; PubMed=3540939;
RA Warrick H.M., de Lozanne A., Levinwand L.A., Spudich J.A.;
RT "Conserved protein domains in a myosin heavy chain gene from Dictyostelium discoideum.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RC STRAIN=AX2;
RX MEDLINE=90353583; PubMed=2387408;
RA Lueck-Vielmeier D., Schleicher M., Grabatin B., Wippler J., Gerisch G.;
RT "Replacement of threonine residues by serine and alanine in a phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
RL FEBS Lett. 269:239-243(1990).
RN [3]
RP PHOSPHORYLATION SITES.
```

```
RX MEDLINE=88112226; PubMed=28281113;
RA Wagle G., Noegel A., Scheel J., Gerisch G.;
RT "Phosphorylation of threonine residues on cloned fragments of the Dictyostelium myosin heavy chain.";
RL FEBS Lett. 227:71-75(1988).
RN [4]
RX X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX MEDLINE=95345066; PubMed=7619795;
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M., Rayment I.;
RT "X-ray structures of the myosin motor domain of Dictyostelium discoideum complexed with MgADP.Befx and MgADP.ALf4-.";
RL Biochemistry 34:8960-8972(1995).
RN [5]
RX X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX MEDLINE=95345067; PubMed=7619796;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the truncated head of Dictyostelium discoideum myosin to 2.7-A resolution.";
RL Biochemistry 34:8973-8981(1995).
RN [6]
RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=96206189; PubMed=8611530;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II).ADP.vanadate complex of the Dictyostelium discoideum myosin motor domain to 1.9-A resolution.";
RL Biochemistry 35:5404-5417(1996).
RN [7]
RX X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX MEDLINE=97452580; PubMed=9305951;
RA Gullick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT "X-ray structures of the MgADP, MgATPgammA, and MgAMPNP complexes of the Dictyostelium discoideum myosin motor domain.";
RL Biochemistry 36:11619-11628(1997).
RN [8]
RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=98070605; PubMed=9405148;
RA Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
RT "X-ray crystal structure and solution fluorescence characterization of Mg.2'(3')-O-(N-methylanthraniloyl) nucleotides bound to the Dictyostelium discoideum myosin motor domain.";
RL J. Mol. Biol. 274:394-407(1997).
CC -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -!- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED.. IT SELF-ASSEMBLES INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL CORTEX.
CC -!- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN BE FURTHER SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES THE ACTIN-ACTIVATED ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1 POSITION (688).
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
```

```
CC EMBL; M14628; AAA33227.1; -
DR PIR; A26655; A26655.
DR PIR; S00250; S00250.
DR PDB; 1MMA; 03-DEC-97.
DR PDB; 1MMD; 17-AUG-96.
DR PDB; 1MMG; 03-DEC-97.
DR PDB; 1MMN; 03-DEC-97.
DR PDB; 1MND; 17-AUG-96.
DR PDB; 1MNE; 17-AUG-96.
DR PDB; 1VOM; 23-DEC-96.
DR PDB; 1LVK; 28-JAN-98.
DR Dictyob; DD01008; mhca.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR004009; myosin_N.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRODOM; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
DR Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
KW Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
FT DOMAIN 1 761
FT MYOSIN HEAD-LIKE.
FT DOMAIN 762 791
FT IQ.
FT DOMAIN 817 2116
FT COILED COIL (POTENTIAL).
FT NP_BIND 179 186
FT ATP.
FT DOMAIN 638 660
FT ACTIN-BINDING.
FT DOMAIN 738 752
FT ACTIN-BINDING.
FT MOD_RES 130 130
FT METHYLATION (DI-) (POTENTIAL).
FT MOD_RES 678 678
FT ALKYLATION (SH-1).
FT MOD_RES 1823 1823
FT PHOSPHORYLATION (BY MHCK).
FT MOD_RES 1833 1833
FT PHOSPHORYLATION (BY MHCK).
FT MOD_RES 2029 2029
FT PHOSPHORYLATION (BY MHCK).
FT SEQUENCE 2116 AA; 243871 MW; 2FC3770BBIIE56A1 CRC64;

Query Match 6.1%; Score 138.5; DB 1; Length 2116;
Best Local Similarity 21.3%; Pred. No. 7.3;
Matches 75; Conservative

Qy 79 LKPLAGVYRSLKKQIEKNIFTNLNLNLSRLKRRKRYFLDV-LESLDMQFKHISNEY 137
Db 1673 IKSLVAEYDEYQLEDEI-----LAKDKLVKAKRALEVELEVRODLEEEEDSR 1723
Qy 138 IIEDSFLLNSEOKNTLLSKYIKESVENDIKFAQEGISYEVKYLAKYKDDLESIKKVI 197
Db 1724 ELEDKRRLLTTEVED-ISKKI---DAEVEQTKL-----DEAKKLTDDVDTLKKQL 1771
Qy 198 KEERKEFPSPPTTPPSPAKTDQKESKFLPFTLNITETLYNNLV-----NKIDDYLYN 251
Db 1772 EDEKKKLNES-----ERAKKLESENEDFLAKLDAEVKNRSRAEKDRKKYKEDLKD 1822
Qy 252 LKAKINDCNVEKDEAHVKITKLSLDAIDDKIDLPKPYDPEATKLLINDTKMDLGLK 311
Db 1823 TKYKLNDEAATKTQTEIGAAL-----EDQIDELRSKLEQEQAKATQADKSKRTLEGEI 1876
Qy 312 LSTGLVQNFPTIISKL-----TEGFODMLNISQHCYKQCPENSGCFRHL---DER 362
Db 1877 DNLRAQTEDEGKIKNRLEKEKRALEGELEELRETVEEAEDSKSBAOSKRLVLELEDA 1936
Qy 363 EECKLLNYKQEGDKCVENPNPTNENNGGCDATCTEEDSGSSRRKITEC 414
Db 1937 RNLQKEIDAKEIAEDAKSNLQREIVEAKGRLE-BESARTNSDRSRKLEAE 1987

RESULT 18
SBCC_CLOAB
ID SBCC_CLOAB STANDARD; PRT; 1163 AA.
AC Q97FK1;
DT 15-JUN-2002 (Rel. 41, Created)
```

15-JUN-2002 (Rel. 41, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
Nuclease sbccD subunit C.
SBCC OR CAC2736.
Clostridium acetobutylicum.
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
NCBI_TaxID=1488;
(1)
SEQUENCE FROM N.A.
STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
MEDLINE=21359325; PubMed=11466286;
Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
Bennett G.N., Koonin E.V., Smith D.R.;
"Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum.";
J. Bacteriol. 183:4823-4838(2001).
-!- FUNCTION: SbccD cleaves DNA hairpin structures. These structures
can inhibit DNA replication and are intermediates in certain DNA
recombination reactions. The complex acts as a 3'->5' double
strand exonuclease that can open hairpins. It also has a 5'
single-strand endonuclease activity (By similarity).
-!- SUBUNIT: Heterodimer of sbcc and sbcd (By similarity).
-!- SIMILARITY: BELONGS TO THE SMC FAMILY. SBCC SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

CC EMBL; AE007771; AA80682.1; -
DR InterPro; IPR003439; ABC_transportr.
KW Hydrolyase; Nuclease; Exonuclease; Endonuclease; DNA replication;
DNA recombination; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 35 42
FT ATP (POTENTIAL).
FT DOMAIN 197 415
FT COILED COIL (POTENTIAL).
FT DOMAIN 446 1003
FT COILED COIL (POTENTIAL).
FT SEQUENCE 1163 AA; 135507 MW; CESF0BD2215D7A92 CRC64;
Query Match 6.0%; Score 136; DB 1; Length 1163;
Best Local Similarity 19.8%; Pred. No. 5;
Matches 85; Conservative
Qy 64 DNI--LSGFENEYDVYIKPLAGVYRSLKKQIEKNIFTNLNL-----NDI-----L 108
Db 677 DNIKLSERSNKIEVEYQK-----KTVEQCEKRIVDLKSLEAEIKERFNAVYTIENL 731
Qy 109 NSRLKKRYK---FDVLDES-----LMQFKH-----ISS 134
Db 732 KAEKIQDFKEMKEILEKERVVEAGEIKDLNRLNIRTEKEQLMDKSRSLKEELSK 791
Qy 135 NEYTIQDSFKLLNS-----EQNTLLSKYIKESVENDIKFAQ-----174
Db 792 NKAELKEDKDIINEKIELKKNVGLDNLVELKEKIEGTIKKIEQVNLCDKKMNEIEDK 851
Qy 175 -----GISYEVKYLAKYD-----DLESIKKVIKEEKFPSPPTTPPSPAKTDQKKE 224
Db 852 YRKCSDEILKIYHSLSSLSKDRKVNIDIKNLKILMBEKF-----NTEKAKENYLN 901
Qy 225 SKFLPFTNIETLYNNLVNKKIDDDYLYNKKAKINDCNVEKDE-----AHVKITKLSLDAID 280
Db 902 DKEINLKSQVEKYKNSLKVNGAVEVLSKLLKNRKLTEERKWIQLQNNRVEKASKAKALQ 961
Qy 281 DKIDLFPNYPDEAIFKLLINDDTKMDLGLKLLSTGLVQNFPTI-----ISKLEEG-KFQ 334
Db 962 ER-----SIKLEEEVKNTIEIKLGLGLKLT--KQLEHKLKSLDLDLEKLFKGRKFV 1011
Qy 335 DMLNISQHCV-----KKOCPENSGCFRHLDERECKCLL-NYKQEGDKCVENPNPTCNEN 389

```
Db 1012 EFVALNQLKVIETASKRLKEITGGNYGLEVDNGRFFIIRDYK----- 1054
QY 390 NGCCDADAT 398
Db 1055 NGCKRDAS 1063

RESULT 19
RBPL_PLAVB
ID RBPL_PLAVB STANDARD; PRT; 2869 AA.
AC Q00798;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Reticulocyte binding protein 1 precursor.
GN RBPL
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites."
RL Cell 69:1213-1226(1992).
CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
HUMAN RETICULOCYTE CELLS.
CC -!- SUBUNIT: HOMODIMER (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Membrane-bound.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL: M88097; AAA29743.1;
KW Malaria; Receptor; Signal; Transmembrane.
FT SIGNAL 1
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCFF CRC64;

Query Match 6.0%; Score 135.5; DB 1; Length 2869;
Best Local Similarity 21.9%; Pred. No. 15;
Matches 79; Conservative 70; Mismatches 136; Indels 75; Gaps 17;

QY 91 KQTEKNFTNLMNDI-----LNSRLKRRKRYFLDVL-----ESDLMQFKHIS 133
Db 2265 RDIKKELYLFQHN-SDISIVEGGVQNMALYDLKNEKREMDLYRNISETKLQWEHST 2323
QY 134 SNEYIETDFKLLNSQKNTLLKSYKIK-----ESVENDIKFAQEGI-----SY 178
Db 2324 DVFKPMIELKHGNETNKNLSLEKEKLLKSVNDHMHSMEAEMIKNGLKYTPESVQNIINI 2383
QY 179 YEKVLAKYKDDLESIKK-----VTKEEKFPSSPTTPPSPAKTDEOKKESKFLPFL 231
Db 2384 YSVIEAEVK-TLEIDRDYGDNTQVIEHKKQFSILIDRTNALMDDIEFKKENNYNLM 2442
QY 232 TNETLYNNLVNKIDYDLINKAKINDCNVKEDEAHVKITKLSLKAIDDKI-DLFKNPY 290
Db 2443 VNTET-----IHRVNDYIEKITNKLVOAKTEYEQI-----LENIKQNDMDLQNIPLKV 2491
QY 291 D-----FEAIKK-----LINDTKKMDLGLKLLSTG-----LVONFPNTTISKLECKFDML 337
```

```
Db 2492 SIIEFYFNKVKKESILNDLYEQE---RLKIGEHLDIEKRNVTEITSSYIEDQKMEMM- 2547
QY 338 NISOHOQVKKQCPENSGCFRHLDERECKCLLAKYKQ--EGDKCVENPPTCNENNGCDA 395
Db 2548 --SKNLEKRSKMMNYTSIYEL-EREANEINRDAKQIKDDDTILNSVLEAAIKRGMDA 2604

RESULT 20
RA50_PYRAB
ID RA50_PYRAB STANDARD; PRT; 880 AA.
AC Q9U2C8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR PAR0812.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae.
OX Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GE5 / Orsay;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
structure and evolution."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
rad50/mrell complex possesses single-strand endonuclease activity
and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mrell by unwinding
and/or repositioning DNA ends into the mrell active site (By
similarity).
CC -!- SUBUNIT: Forms a complex with mrell (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL: AJ248286; CAB50131.1;
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02463; SMC_N; 1.
DR ProDom: PD000006; ABC_transportr; 1.
DR SMART: SM00382; AAA; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 30 37 ATP (BY SIMILARITY).
FT DOMAIN 144 745 COILED COIL (POTENTIAL).
SQ SEQUENCE 880 AA; 103970 MW; FDB177EC7E026479 CRC64;

Query Match 5.9%; Score 135; DB 1; Length 880;
Best Local Similarity 24.0%; Pred. No. 4; 1;
Matches 79; Conservative 48; Mismatches 96; Indels 106; Gaps 13;

QY 79 LKPLAGVYRSLKQKQIEKNFTNLMNDILNSRLKRRKRYFLDVLSDLMQFKH----- 131
Db 237 LESIKGKISLKIQVEK-----LKGRK---KGLEEKIVQIERSIEEKA 277
QY 132 -ISSNEVLIEDSKLLNSQKNTLLKSYKIKESVENDIKFAQEGISYKYLAKYKDDL 190
Db 278 KISELEEIVADIPKLOKEKEKRYKLGKGFDEYES-----KLRLEKELSKWSEL 327
QY 191 ESTKKVKE-EKEKFPSSPTTPPSPAKTDEOKKESKFLPFLTNETLYNNL---VNKID 246
Db 328 KATEEVKEKEKK-----ERAEIREK-----LSEIERLEELKPYVELE 369
QY 247 DYLINKAKINDCNVKEDEAHVK-----ITKLSLKAIDDKIDLFPKNPYDFAIKKLI 299
```

Db 370 D-----AKOVQKQIRLAKRLKGLSPGVEIKLESLEKTERTEIE-----EAIKEI- 414
 Qy 300 NDDTKKMLGKLLTGLVQNPNTIISKIEKQDMLNISOHQVKKQCPNSCCPRHL 359
 Db 415 -----TTRIGQMEQKERNKMAIELRAKAKGCPV---CGREL 449
 Qy 360 DERECKCLLNYKQEGDKVNPNTCNE 388
 Db 450 TEHHKELMERYTLEIKKIEELKRTTEE 478

RESULT 21

USO1_YEAST STANDARD; PRT; 1790 AA.
 AC P25386;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Intracellular protein transport protein USO1.
 GN USO1 OR INT1 OR YDL058W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=X2180-1A;
 RC MEDLINE=91185402; PubMed=2010462;
 RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
 RA Yamasaki M.;
 RT "A cytoskeleton-related gene, usol, is required for intracellular
 RT protein transport in Saccharomyces cerevisiae.";
 RL J. Cell Biol. 113:245-260(1991).
 RN [2]
 RP SEQUENCE OF 782-1790 FROM N.A.
 RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
 RA Kendrick K.E.;
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-8 FROM N.A.
 RA Bai Y., Symington L.S.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
 CC COMPLEX.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
 CC MEMBRANES. PROBABLY PRESENT ON VESICLES. OPERATIONAL BETWEEN THE
 CC ER AND THE GOLGI COMPLEX.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
 CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
 CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
 CC -1- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X54378; CAA38253.1; -;
 DR EMBL; L03188; AAB00143.1; -;
 DR EMBL; U53668; AAB66659.1; -;
 DR PIR; A38455; A38455.
 DR SGD; S0002216; USO1.
 DR InterPro; IPR002017; Spectrin.
 KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
 FT DOMAIN 1 724 GLOBULAR HEAD.
 FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
 FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
 FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
 FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).

FT CONFLICT 847 847 G -> E (IN REF. 2).
 FT CONFLICT 924 924 E -> K (IN REF. 2).
 FT CONFLICT 1253 1253 V -> I (IN REF. 2).
 FT CONFLICT 1319 1319 I -> V (IN REF. 2).
 FT CONFLICT 1461 1461 N -> S (IN REF. 2).
 FT CONFLICT 1581 1581 G -> S (IN REF. 2).
 FT CONFLICT 1600 1600 I -> V (IN REF. 2).
 FT CONFLICT 1661 1661 R -> S (IN REF. 2).
 FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
 SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;
 Query Match 5.9%; Score 134; DB 1; Length 1790;
 Best Local Similarity 21.2%; Pred. No. 10;
 Matches 99; Conservative 77; Mismatches 191; Indels 100; Gaps 19;
 Qy 20 ETAAKFERQHMDSPDLGTDODDKAMADIGSIGRGR-----TMAISVTMDNILLSGFENE 73
 Db 1165 EEQIANKEQYNEISQNLDEITSTQOENESIKKKKNDELEGEVAMKSTSEEQSNLKKE 1224
 Qy 74 YDVIYLPAGVYRSLKQIEKNIFTNLDNLNLDILNSRLKKRVFLDVLESDLMQFKHIS 133
 Db 1225 IDALNLO-----IKELKKNETNEASLLESIKSVSEIVKIKEL-----QDECNFKKE 1273
 Qy 134 SNEYIIEFSKLNSEOKNT-----LLSKYKIKESVE---NDIKPAQEGISYYEKVLAKY 186
 Db 1274 VSE--LED--KLKASEDKNSKYLELQKESIKKELDAKTTTELKIQLEKITNLSKAKEKS 1329
 Qy 187 KDDLESIKKVIKEK-----EKFPSPPTTPSPAKTDQKKESKEL-PFLTNIELTY 238
 Db 1330 ESELRLKTTSEERKNAEOLKLNKNE-----IQIKNAFAFERKLNLSGSSSTITQEQY 1383
 Qy 239 NNLVKNIDDIYLI-----NLKAKINDCNVERDEAHVKIT-----KLSDLKADDDKIDL 285
 Db 1384 SEKINTLEDELIRLQENENELAKKEID-NTRSELEKVSLSNDELLEKEKQNTIKSLQDEILS 1442
 Qy 286 FK-----NPDYFAIK-----KLNIDTKKMDLGLKLL 312
 Db 1443 YKDKITRNDKLLSIEDNRKRDLESLEQLRAAESKAKVEEGLKKLEESKKALEK 1502
 Qy 313 STGLVQNPNTIISKIEGKFDMLNISQH-----QCVKKQCPNSGCPRHLDDEBECKCL 368
 Db 1503 SKEMMKLESTIESNETELK-SSMETIRKSDKLEQSKSAEEDIKNLOH-----EKSDEL 1556
 Qy 369 LNYKQEGDKVNPNTCN-ENNGGCDADATCTEEDSGSSRKKITCE 414
 Db 1557 ISRINESEKQIEELKRLRIEAKSGSELETVKQELNNAQEKIRINAE 1603

RESULT 22

MYS1_YEAST STANDARD; PRT; 1928 AA.
 ID AC P08964;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin-1 isoform (Type II myosin).
 GN MYO1 OR YHR023W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RC MEDLINE=91088308; PubMed=2263482;
 RA Sweeney F.P., Watts F.Z., Pocklington M.J., Orr E.;
 RT "The MYO1 gene from Saccharomyces cerevisiae: its complete nucleotide
 RT sequence.";
 RL Nucleic Acids Res. 18:7147-7147(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RC MEDLINE=94378003; PubMed=8091229;

Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
 Du Z., Favella A., Fulton L., Gattung S., Geisel C., Kirsten J.,
 Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
 Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
 Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
 Vignati D., Wilcox L., Wohlman P., Waterston K., Wilson R.,
 Vaudin M.;
 *Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 VIII.*;
 Science 265:2077-2082(1994).
 [3]
 RP SEQUENCE OF 1-760 FROM N.A.
 RC STRAIN-5288c;
 RX MEDLINE=88111539; PubMed=3322809;
 RA Watts F.Z., Shields G., Orr E.;
 *The yeast MYO1 gene encoding a myosin-like protein required for cell
 division.*;
 RT EMBO J. 6:3499-3505(1987).
 RL Science 265:2077-2082(1994).
 CC -|- FUNCTION: REQUIRED FOR CELL DIVISION.
 CC -|- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -|- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch)
 CC -----
 CC EMBL: X53947; CAA37894.1; -;
 DR EMBL: X06187; CAA29550.1; -;
 DR EMBL: U01399; AAB68872.1; -;
 DR PIR: S05806; S05806.
 DR PIR: S12323; S12323.
 DR PIR: S46773; S46773.
 DR HSP: P08799; IAMD.
 DR SGD: S0001065; MYO1.
 DR InterPro: IPR000048; IQ_region.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00063; myosin_head; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR ProDom: PD000355; myosin_head; 1.
 DR SMART: SM00015; IQ; 1.
 DR SMART: SM00242; MYSC; 1.
 DR PROSITE: PS50096; IQ; 1.
 KW Myosin; Actin-binding; ATP-binding; Coiled coil; Alkylation.
 FT DOMAIN 1 793
 FT DOMAIN 794 823
 FT NP_BIND 856 1911
 FT NP_BIND 180 187
 FT DOMAIN 460 529
 FT MOD_RES 703 703
 FT CONFLICT 36 36
 FT CONFLICT 46 46
 FT CONFLICT 59 59
 FT CONFLICT 86 86
 FT CONFLICT 330 330
 FT CONFLICT 343 343
 FT CONFLICT 421 426
 FT CONFLICT 513 515
 FT CONFLICT 529 535
 FT CONFLICT 541 541
 FT CONFLICT 550 551
 FT CONFLICT 573 573
 FT CONFLICT 582 582
 FT CONFLICT 588 599
 FT CONFLICT 599 599
 FT CONFLICT 627 632
 FT CONFLICT 695 695
 FT CONFLICT 736 742
 FT CONFLICT 756 756
 FT CONFLICT 773 784

FT CONFLICT 793 794
 FT CONFLICT 896 896
 FT CONFLICT 900 900
 FT CONFLICT 906 906
 FT CONFLICT 911 911
 FT CONFLICT 915 930
 FT CONFLICT 934 939
 FT CONFLICT 951 953
 FT CONFLICT 955 958
 FT CONFLICT 1002 1002
 FT CONFLICT 1049 1049
 FT CONFLICT 1056 1056
 FT CONFLICT 1060 1060
 FT CONFLICT 1085 1085
 FT CONFLICT 1123 1123
 FT CONFLICT 1133 1133
 FT CONFLICT 1144 1146
 FT CONFLICT 1159 1168
 FT CONFLICT 1179 1181
 FT CONFLICT 1184 1185
 FT CONFLICT 1198 1204
 FT CONFLICT 1224 1224
 FT CONFLICT 1228 1228
 FT CONFLICT 1253 1253
 FT CONFLICT 1311 1323
 FT CONFLICT 1400 1400
 FT CONFLICT 1454 1554
 FT CONFLICT 1568 1568
 FT CONFLICT 1630 1646
 FT CONFLICT 1698 1704
 FT CONFLICT 1725 1737
 FT CONFLICT 1754 1757
 FT CONFLICT 1777 1777
 FT CONFLICT 1788 1788
 FT CONFLICT 1825 1825
 FT CONFLICT 1882 1882
 FT CONFLICT 1902 1904
 SQ SEQUENCE 1928 AA; 223634 MW; 6F54C761F43DC9F CRC64;
 Query Match 5.9%; Score 134; DB 1; Length 1928;
 Best Local Similarity 23.4%; Pred.No.11;
 Matches 96; Conservative 70; Mismatches 133; Indels 112; Gaps 25;
 QY 17 GKMT--AAAKFERQHMDSPDLGTDGDDKAMADIGSIEGRGTMAISVTMDNLSGFENE 73
 DB 963 GLOETTRERATLEKLSKNNEL-----IKQISDLNCDISKQSSOSLTKESKLK-LENE 1016
 QY 74 YDVIYKPLAGVYRSLKQIEK-----NIEFTNLNLDINLSRLKRRKYFL 119
 DB 1017 -----IKRLKDVINSKEEBIKSPNDKLSSEEDLDIKLVTEKNCN-IAMSLQS---L 1066
 QY 120 DVLESIMOFKHISSENYIEDSFK-----LLNSEQN---TLKSKYIKESVENDIK-- 170
 DB 1067 VTENSOLR-----SKNE-----NFKKEKAALNNQLNKSELEK---MKEKIDNHKKEL 1112
 QY 171 --PAQ--EGISYVEKVLAKYD-----DLESIKKVIKEKEKFFSPPTTPPSPKTD 219
 DB 1113 AFSKORDDAVSEHGKITAEKRETRIQLETKYKSNYKIKEEYNFORE-----TKEQ 1164
 QY 220 EQKESKELPFLNIEFTLYNNLVNKIDYLL---INLAKIND-CNVEKDEAHVKITKLS 275
 DB 1165 EQKRNRL-----VESLNDKIKLEARLSQELISLNOYLKRNISGNSVETNISSTRST 1218
 QY 276 LKAIID--DKIDLPKNPYDF-----EAIKILIND-DYFKDMLGKLSTGLVONFPNTIISK 327

CC	EMBL; U05820; AAA17416.1; -
DR	EMBL; DS0617; EAA09270.1; -
DR	PIR; S48530; S48530.
DR	SGD; S0001927; SMC2.
DR	InterPro; IPR003405; SMC_C.
DR	InterPro; IPR003395; SMC_N.
DR	Pfam; PF02463; SMC_N; 1.
DR	Pfam; PF02483; SMC_C; 1.
DR	ProDom; PD000006; ABC transport; 1.

```

RESULT 24
RA50_AQUAE
ID RA50_AQUAE STANDARD; PRT; 978 AA.
OC O67124;
DT DT 16-OCT-2001 (Rel. 40, Created)
DT DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DE Probable DNA double-strand break repair rad50 ATPase.
OS RAD50 OR AQ.1006.
GN Aquifex aeolicus.
OC Bacteri; Aquificae; Aquificae (class); Aquificaceae; Aquificaceae;
OC Aquifex.
OC NCBI_TaxID=63363;
RN RN [1]
RP RP SEQUENCE FROM N.A.
RC RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=95373320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aufay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RA "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RT Nature 392:353-358(1998) .
RL
CC CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC CC rad50/mre11 complex possesses single-strand endonuclease activity.

```



```
FT DOMAIN 1 828 MYOSIN HEAD-LIKE.
FT DOMAIN -829 2104 COILED COIL (POTENTIAL).
FT DOMAIN 646 660 ACTIN-BINDING (POTENTIAL).
FT NP_BIND 182 189 ATP (POTENTIAL).
FT MOD_RES 686 686 ALKYLATION (BY SIMILARITY).
FT CONFLICT 1193 1193 D -> G (IN REF. 2).
FT CONFLICT 1304 1304 E -> G (IN REF. 2).
FT CONFLICT 1344 1344 E -> K (IN REF. 2).
FT CONFLICT 1420 1420 G -> D (IN REF. 2).
SQ SEQUENCE 2104 AA; 242570 MW; 3A0548594028D258 CRC64;

Query Match 5.88; Score 133; DB 1; Length 2104;
Best Local Similarity 21.58; Pred. No. 14;
Matches 73; Conservative 55; Mismatches 122; Indels 90; Gaps 11;

QY 77 IYLPLAGVYRLSKKQTEKNIFTN-----NLNLSRLKRRKRYFLDVLDESLLMOPKH 131
Db 823 LHLAPLLGTTOTDEYLRKRDALINLNQLOESTKEVANELTITKERVLO-LTNDLQEEQA 881

QY 132 ISSNEYI-----IEDFKLLNSEQNTLLKSYK-----IKESVENDIKFAQEGISY 178
Db 882 LAHEKDILVERANSRVEVHERLSLENOQVTIADEKVEFLYAEKOSTEEDLANKOTETSY 941

QY 179 YEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTPSPAKTDEQKESKFLPLNIETLY 238
Db 942 LSDLSSTLEKKLSISK-----DEQTISSKY-----KELE 971

QY 239 NLNVNKIDDDY-----LNLKAKINDCNVEKDEAHVKITKLSDLKADDKIDLFKNPYDF 292
Db 972 KYLNIMADYQHSOHLNSLEKAINELNRELNEKMLRL-----DDELLLKQORSYDT 1025

QY 293 EAIKKLINDTKKMKGLKLS--TGLVQNPNTTIISKLEGFQDMNLNISOHCVKKQCP 350
Db 1026 KVQELRENASLKQCRYESQLASYSKYSET-----ESELNKEAEVLVFO-- 1073

QY 351 ENSGCFRHLDERECKLLNKKQEGDKVENPNPTCENN 390
Db 1074 -----KEITEYRQDLHRAQFQNPETHNIND 1098

RESULT 26
RA50_SULAC
ID RA50_SULAC STANDARD; PRT; 886 AA.
AC O33600;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA double-strand break repair rad50 Atpase.
GN RAD50.
OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2285;
RN [1]
SEQUENCE FROM N.A. / NCIB 11770 / DSM 639;
RC STRAIN=ATCC 33909 / PubMed=9211741;
RA Elie C., Baucher M.F., Fondrat C., Forterre P.;
RT "A protein related to eucaryal and bacterial DNA-motor proteins in the hyperthermophilic archaeon Sulfolobus acidocaldarius.";
RL J. Mol. Evol. 45:107-114(1997).
CC -/- FUNCTION: Involved in DNA double-strand break repair (DSBR). The rad50/mrell complex possesses single-strand endonuclease activity and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mrell by unwinding and/or repositioning DNA ends into the mrell active site (By similarity).
CC -/- SUBUNIT: Forms a complex with mrell (By similarity).
CC -/- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y10687; CAA71688.1; -
CC InterPro; IPR003439; ABC_transportr.
CC InterPro; IPR003395; SMC_N.
CC Pfam; PF02463; SMC_N; 1.
CC ProDom; PD000006; ABC_transportr; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil.
NP_BIND 30 37 ATP (BY SIMILARITY).
FT DOMAIN 174 727 COILED COIL (POTENTIAL).
SQ SEQUENCE 886 AA; 103857 MW; 0390AE1403194104 CRC64;

Query Match 5.88; Score 132.5; DB 1; Length 886;
Best Local Similarity 23.68; Pred. No. 5.5;
Matches 89; Conservative 67; Mismatches 142; Indels 79; Gaps 18;

QY 62 TMDNILSGFENEYDVIYLLKPLAGVYRSLKQIE-----KNIFTNLSNLDILN----- 109
Db 186 SIKDILKREAEID-----RLKKEIEIKVKLENIEREAKEKEDELNOYNTEF 233

QY 110 SRLKRRYFLDVLSDL-----MQFKHSSNE---YIIEDSFKLL--NSEQKNT 153
Db 234 NRIKEIKVQVDILSGELSVNKNKIEETALRLKDFPEKEKRYNKIETEVKELDENREKINT 293

QY 154 L-----LKSQYKTESVENDIKFAQEGISYVEKVLAKYK--DDLESKKVKIKEKE 202
Db 294 ISSFKSLVILQIDSLKQINNVENDLKRKKEKLRKKELEKEKEQVEIEKKEKKEEKEK 353

QY 203 KFPSPPTTPSPAKTDEQKESKFLPLNIETLYNNL-----VNKIDYLLNLKAKI 256
Db 354 QYEEIKRLTVLVKNIERQKNIEKLVV--DTQDLENKKIKDVSDRINQIDNELKGLDDR 412

QY 257 NDCNVEKDEAHVKITKLSDLKAI--DDKIDLFKNPYDFAIKKLLNDDTKKMKGLKLISTG 315
Db 413 GDLNGRKEQT-LKI--YNNLSIEDDRCPICGRPLDSEH-KAKIRIEIKVOLLLELNKQIT 468

QY 316 LVQNPNTTIISKLEGFQDMNLNISOHCVKKQCPENSGCF-----RHUDE-----RE 363
Db 469 ALQARINSLIKEREE--LEATRNKLQLELQKRS--KEGIYEAKLKLQRLLEEKKNLQON 524

QY 364 ECKCLNKKQEGDKCQE 380
Db 525 EILSLLSYHQEFNTAE 541

RESULT 27.
RA50_PYRFU
ID RA50_PYRFU STANDARD; PRT; 882 AA.
AC P58301;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA double-strand break repair rad50 Atpase.
GN RAD50 OR PF1167.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RX MEDLINE=20485553; PubMed=11029422;
RA Hopfner K.-P., Karcher A., Shin D., Fairley C., Tainer J.A., Carney J.P.;
RT "Mrell and Rad50 from Pyrococcus furiosus: cloning and biochemical characterization reveal an evolutionarily conserved multiprotein machine.";
RL J. Bacteriol. 182:6036-6041(2000).
RN [2]
SEQUENCE FROM N.A.
```

RC STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 1-149.
RX STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RX MEDLINE=2034838; PubMed=10892749;
RA Hopfner K.-P., Karcher A., Shin D.S., Craig L., Arthur L.M.,
RA Carney J.P., Tainer J.A.;
RT "Structural biology of Rad50 ATPase: ATP-driven conformational
RT control in DNA double-strand break repair and the ABC-ATPase
RT superfamily.";
RL Cell 101:789-800(2000).
CC -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site.
CC -1- SUBUNIT: Forms a complex with mre11.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE010225; AAL81291.1; -
DR PDB; 1F2T; 02-AUG-00.
DR PDB; 1F2U; 02-AUG-00.
DR InterPro: IPR003439; ABC transportr.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02463; SMC_N; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; 3D-structure;
FT Complete proteome.
FT NP_BIND 30 37 ATP.
FT DOMAIN 148 744 COILED COIL (POTENTIAL).
SQ SEQUENCE 882 AA; 103839 MW; 3ADCBD250382A99E CRC64;
Query Match 5.8%; Score 132; DB 1; Length 882;
Best Local Similarity 23.4%; Pred. NO. 5.8;
Matches 78; Conservative 73; Mismatches 101; Indels 82; Gaps 21;
QY 52 EGRGTWAI5VTMDNLSGPNEDVLYLPLAGVYRSLKKQIEKNFTFNLDILNSR 111
Db 186 EYRDIILARTENIEELIK--ENEQELI-----OVLOEISK-----TEEVLP SK 225
QY 112 LKRRYFLDVLSDLMQFHHISSNEYIIEDSPKLLNSEQKNTLLSKYIKESVENDIKF 171
Db 226 RSK-----VDMLRVELRL-----ETKVEIENSER---LLEKRGDKRLEGRKN 269
QY 172 AQGISYIEKVLAKYKDDLESIKV--IKEKEKFPSPPTPPSPAKTDEQKESKFLP 229
Db 270 TEE---YLEKLKEKEKELEEQVEITSIKKVDAYLAL-----KEFKNEYLQKKYKIEK 320
QY 230 FLTNIEFLYNNLVNKIDY-----LINKAKIND--CNVEKD-EAHVKI-TKLSDL 276
Db 321 ELTRVBEILNEIQKLEELNEKESEKEKLEKLEKLEKLEKLEKLEKLEKLEKLEK 380
QY 277 KATDDKIDLFKNPYDFEAKKLLND-DTKKMDLGLKLLSTGLVQNFNPNTIISKL--TEGKF 333
Db 381 RQKLEKLG-DKSPED---IKKLEELTKT-----TIEERNEITQRIKELNKNI 427
QY 334 QDLNL-ISOHQCVKQCPENSGCFRHL--DEREE 364
Db 428 GDLKTAIEELKAKGKCPV---CGRELTDEHREE 458
RESULT 28
BRC2_MOUSE

ID BRC2_MOUSE STANDARD; PRT; 3329 AA.
AC P97929; P97383; O35922;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Breast cancer type 2 susceptibility protein.
GN BRC2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=97217789; PubMed=9063750;
RA Connor F., Smith A., Wooster R., Stratton M., Dixon A., Campbell E.,
RA Tait T.M., Freeman T., Ashworth A.;
RT "Cloning, chromosomal mapping and expression pattern of the mouse
RT Brc2 gene.";
RL Hum. Mol. Genet. 6:291-300(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=97237041; PubMed=9119389;
RA Sharan S.K., Bradley A.;
RT "Murine Brc2: sequence, map position, and expression pattern.";
RL Genomics 40:234-241(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=97384941; PubMed=9242436;
RA McAllister K.A., Haugen-Strano A., Hagevik S., Brownlee H.A.,
RA Collins N.K., Futreal P.A., Bennett L.M., Wiseman R.W.;
RT "Characterization of the rat and mouse homologues of the BRC2 breast
RT cancer susceptibility gene.";
RL Cancer Res. 57:3121-3125(1997).
RN [4]
RP SEQUENCE OF 18-200 FROM N.A.
RX MEDLINE=97075121; PubMed=8917547;
RA Rajan J.V., Wang M., Marquis S.T., Chodosh L.A.;
RT "Brc2 is coordinately regulated with Brcal during proliferation and
RT differentiation in mammary epithelial cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:13078-13083(1996).
RN [5]
RP SEQUENCE OF 569-625 FROM N.A.
RX MEDLINE=97341126; PubMed=9196008;
RA McAllister K.A., Ramchandran S., Haugen-Strano A., Fiedorek F.T. Jr.,
RA Wiseman R.W.;
RT "Genetic mapping of the Brc2 breast cancer susceptibility gene on
RT mouse chromosome 5.";
CC Mamm. Genome 8:540-541(1997).
CC -1- FUNCTION: MAY PARTICIPATE IN A PATHWAY ASSOCIATED WITH THE
CC ACTIVATION OF DOUBLE-STRAND BREAK REPAIR AND/OR HOMOLOGOUS
CC RECOMBINATION (BY SIMILARITY).
CC -1- SUBUNIT: INTERACTS WITH RAD51 (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST EXPRESSION IN
CC CEREBELLUM, TESTIS, ILEUM, APPENDIX, EPIDYMISS, OVARY AND MAMMARY
CC GLAND. NO EXPRESSION IN LUNG.
CC -1- DEVELOPMENTAL STAGE: IN THE MAMMARY GLAND, EXPRESSION INCREASES
CC DRAMATICALLY DURING PREGNANCY.
CC -1- SIMILARITY: CONTAINS 7 BRC2 REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U82270; AAB48306.1; -
DR EMBL; U72947; AAB40720.1; -
DR EMBL; U65594; AAC23702.1; -


```
Db 333 DOOSEYEQHASFQYRLKLEQLESSKANWDSIEYALN-----TKIVN----- 374
QY 312 LSTGLVQNPNTIISKL-IEGKFQDMLNISO 341
Db 375 -----LENRFSTWKEKNDIEEKYTALRSSE 401

RESULT 30
SMC1_YEAST
ID SMC1_YEAST STANDARD; PRT; 1225 AA.
AC P32908;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chromosome segregation protein SMC1 (DA-box protein SMC1).
GN SMC1 OR CHL10 OR YFL008W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94103320; PubMed=8276886;
RA Strunniakov A.V., Lariouov V.L., Koshland D.;
RT "SMC1: an essential yeast gene encoding a putative head-rod-tail
RT protein is required for nuclear division and defines a new ubiquitous
RT protein family.";
RL J. Cell Biol. 123:1635-1648(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH SMC2 OR OLIGOMERS.
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L06062; AAA16595.1; -;
DR EMBL; D50617; BAA09230.1; -;
DR PIR; A49464; A49464.
DR PIR; S41804; S41804.
DR SGD; S0001886; SMC1.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02463; SMC_N; 1.
DR Pfam; PF02483; SMC_C; 1.
KW Mitosis; ATP-binding; Coiled coil; Nuclear protein.
FT NP_BIND 33 40
FT DOMAIN 173 489
FT COILED COIL (POTENTIAL).
FT DOMAIN 679 1063
FT COILED COIL (POTENTIAL).
FT DOMAIN 1057 1061
FT NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 1137 1164
FT ALA/ASP-RICH (DA-BOX).
FT MUTAGEN 173 173
FT S->L: IN TS MUTANT SMC1-2.
FT MUTAGEN 458 458
FT N->D: IN TS MUTANT SMC1-1.
SQ SEQUENCE 1225 AA; 141279 MW; B504017AA0ECC8C CRC64;
```

```
Query Match 5.8%; Score 131.5; DB 1; Length 1225;
Best Local Similarity 20.9%; Pred. No. 9;
Matches 93; Conservative 72; Mismatches 126; Indels 155; Gaps 23;

QY 38 TDDDDKAMADIGSIEGR-----CTMAISVTMDNILSGFENEYDVII--LKPLAGVY 86
Db 253 TDKLSALNSEISLKGKINNEKSLQSKSFVKESAVISKQSKLDYIFKDKREKLVSDL 312
QY 87 RSLK-----KOJEKNFTFNLDNLNLSRLKRYFLDVLSDLMQFKHISNE 136
Db 313 RLKVPQQAAGKRISHIEKRI-----ESLQKDLQORQTYVERFET---OLKVVTRSK 361
QY 137 YIIE-----DSFKLLNSEQNTLLSKYIKESVENDIKFAQEGISYVEKVLAKY 186
Db 362 EAFEEIKQSARNYDKFKLNEND-----LKYINCLHE-----KYLTEGGSILKEIAVL 410
QY 187 KDOLESKVKIKKEKEKFPSPPTTPPSPAKTDEQKESKFLPLTNIETLYNNLVNKID 246
Db 411 NND-----KREIQBELEFRNK-----RADISKRRITEELSITG-----EKLD 447
QY 247 DYLLINKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKN-PYD----- 291
Db 448 TQLNDLRSVLSN-----EKALHTE---RUHLKLLQSDIESANNQYDNLFKRLRETLVKIDD 501
QY 292 -----FEAKKLIND--DTKKDMLGKLLSTGLVQNPFTII 325
Db 502 LSAHQRETMKEREKLRENIAMLKRRFFGKGLVHDLCHPKKEKYGLAVSTILGKNFDSVIV 561
QY 326 SKLIEGKFQDMLNISQHCV---KKQCPENSGGCFRHLDERE-ECKCL----- 368
Db 562 -----ENLTVAQ-ECTAFLLKKQ-RAGTASFIPDLTETELPTLSLPDSQDYIILSN 610
QY 369 -LNVKQEGDKCQE---NPNPTCENN 390
Db 611 AIDYEPYERAMQYVCGDSIICNTLN 636

RESULT 31
RBP2_PLAVB
ID RBP2_PLAVB STANDARD; PRT; 1251 AA.
AC Q00759;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Reticulocyte binding protein 2 (Fragment).
GN RBP2.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT merozoites.";
RL Cell 69:1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -1- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M88098; AAA29744.1; -;
KW Malaria; Receptor; Membrane.
FT NON_TER 1251 1251
FT NON_SEQUENCE 1251 AA; 143741 MW; 54BA51C7A04AC572 CRC64;
```


RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Cocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon *Archaeoglobus fulgidus*.";
RL Nature 390:364-370(1997).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001032; AAB90211.1; -;
CC TIGR; AF1032; -;
CC InterPro; IPR003439; ABC_transportr.
CC InterPro; IPR003395; SMC_N.
CC InterPro; IPR02017; Spectrin.
CC Pfam; PF02463; SMC_N; 1.
CC DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
CC FT DOMAIN 31 38 ATP (BY SIMILARITY).
CC FT DOMAIN 148 728 COILED COIL (POTENTIAL).
CC FT SEQUENCE 886 AA; 103633 MW; D35641D499A8B58 CRC64;
SQ
Query Match 5.7%; Score 130; DB:1; Length 886;
Best Local Similarity 21.6%; Pred. No. 7.4;
Matches 85; Conservative 63; Mismatches 128; Indels 118; Gaps 18;
QY 47 DIGSI-----EGRGTMAISVTMDNILSGFENEVDVYLKPLAGVYSLKKQIEKNITFNL 102
Db 130 EIDSIIRDDSEIRIIRQIR---IEDYENAW-----KNLGAVIRMLEREKER----- 174
QY 103 NLNDILN--SRLLKKRYFLDVLESDLMOQFKHISSENYIEDSFKLLNSEQKN----- 152
Db 175 -LKEFLSQEQLKQK-----EKKAEIERISEEKSIESLEKLESEVRNLESRLKEL 227
QY 153 ----TLKSYKYIKESVENDIKFAQBGISYEVKVLAKYKDDLESIKKVIKEKEKPPSP 208
Db 228 EEHKSRLSESLKQESSVLQEVGRGLEKLEKLEKQKVEVERIEDLEKAKVEKELPKAE 287
QY 209 P-----TTPSPAKTDQKE-----SKFLPFLNIETLYNN 240
Db 288 RVSILEKLLSEINQALRDVEKREGDLTREAGTQALQKAAEENSDNSEITRIEELERE 347
QY 241 L-----VNKIDYLNILKAKINDCNVKEDEAHVKITKLSDL--KAIDDKIDLF 286
Db 348 LERFEKSHRLLETLPKPDQMOGIAKLEKNLTPD-----KVEKMYDLLSKAKEEKEI- 402
QY 287 KNPYDFEAIKLLINDPTDKMLGKLSLSTGLVQNPFTIISKLETKEGFQDLNISOHCYK 346
Db 403 ----TEKLKLLI---AKSS-----LKTGAQ-----LKKAVE-----ELKSAE 434
QY 347 KQCPENSGCFRHLDEREECKLLNYYKQEGDKCVC 380
Db 435 RTCPV---CGRELDEEHRKNIMAETREMKRTAE 465
RESULT 34
SMC3_YEAST
ID SMC3_YEAST STANDARD; PRT; 1230 AA.

AC P47037;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Chromosome segregation protein SMC3 (DA-box protein SMC3).
GN SMC3 OR YJL074C OR J1049.
OS Saccharomycetes cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WB303;
RX MEDLINE=97474309; PubMed=9335333;
RA Michaelis C., Ciosk R., Nasmyth K.;
RT "Cohesins: chromosomal proteins that prevent premature separation of
RT sister chromatids.";
RL Cell 91:35-45(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC Rose M., Koetter P., Entian K.D.;
RA Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RA Sor F.J.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y14278; CAA74655.1; -;
CC EMBL; Z49349; CAA89366.1; -;
CC EMBL; X88851; CAA61313.1; -;
CC SGD; S0003610; SMC3.
CC InterPro; IPR003439; ABC_transportr.
CC InterPro; IPR003405; SMC_C.
CC InterPro; IPR003395; SMC_N.
CC Pfam; PF02463; SMC_N; 1.
CC Pfam; PF02483; SMC_C; 1.
CC Mitosis; ATP-binding; Coiled coil; Nuclear protein.
CC NP_BIND 32 39 ATP (POTENTIAL).
CC FT DOMAIN 172 482 COILED COIL (POTENTIAL).
CC FT DOMAIN 685 1041 COILED COIL (POTENTIAL).
SQ SEQUENCE 1230 AA; 141336 MW; B152D88F7780341F CRC64;
Query Match 5.7%; Score 130; DB:1; Length 1230;
Best Local Similarity 19.6%; Pred. No. 11;
Matches 92; Conservative 83; Mismatches 169; Indels 126; Gaps 17;
QY 11 LVPRGCMKETAATAAKFERQHMDSPDLGTD-DOKADIGSIEGRGTMAISVT-----MD 64
Db 145 IVPQGIKVALTNKDKERQLLEDVVGAKSFVKLAKSKKMEETQKKQINKEMGELN 204
QY 65 NILSGFENEVDVYLLKPLAGVYSLKKQIEK-----NIFTNL---NLNDILNRLK 113
Db 205 SKLSEMEQE-----RKELEKYNELERNRRIYQFTLYDRELNEVINOMER 248
QY 114 KRKYFLDVLESDLMOQFKHISSENYIEDSFKLLNSEQKNITLKKYKIKE----- 163
Db 249 LDGDYNTVYSSEQYIOELDKREDMDIQVSKLSSEASLUKIKNATDLOQAKLRESEISQ 308

Qy 164 ---SVENDIKFAQBGISYEVKVLAKYDDLESIKKVIKEKEKFPSPPTTPPSAKTDE 220
 Db 309 KLTNNVNIKQVQOQIESNEQRLSDSATLKEIKSIIEQRKOKL-----SKILPRYOEL 362
 Qy 221 QKESKELPFLTNIETLYNNLVNKIDDY-----LINLKAINDCNVEKD 264
 Db 363 TKEEAMVKLQASQOQKORDILKKEGYARKSKDERDTWTHSEIEELKSSIONLELES 422
 Qy 265 EAHVKITKL-SDLKAIDDKIDLFKNPYDFAEIKKLINDDTKKDKMLGLKLLSTGLVQNPNT 323
 Db 423 QLQMDRTSLRKQYSAIDEIE-----ELIDSDINGPDTK-----GOLEDDESE 464
 Qy 324 II---SKLIEG-----KFDML-----NISQHCQVKKOCPPSPSCFPHLEB-- 361
 Db 465 LIHLKQKLSLDRKELWRKEQKLOTVLETLFLSDVQNOQ-----RNVNETM 511
 Qy 362 -REECKLLNKGQEGDKVENPNP---TCNENNGCCADATCTEEDSGSS 407
 Db 512 SRSLANGLIINVKTEKLIKSPESVFTGLGELIKVNDKYKTCAEVIGNS 561

RESULT 35
 HMW2_MYCGE STANDARD; PRT; 1805 AA.
 AC P47460;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cytadherence high molecular weight protein 2 (Cytadherence accessory protein 2)
 DE HMW2 OR MG218.
 GN Mycoplasma genitalium.
 OS Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 33530 / G-37;
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fleischmann J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
 RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.F., Dougherty B.A., Bost K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RA "The minimal gene complement of Mycoplasma genitalium."
 RL Science 270:397-403(1995).
 RN [2]
 RP SEQUENCE OF 557-659 FROM N.A.
 RC STRAIN-ATCC 33530 / G-37;
 RA MEDLINE-96026346; PubMed-7569993;
 RA Peterson S.N., Hu P.-C., Bost K.F., Hutchison C.A. III;
 RA "A survey of the Mycoplasma genitalium genome by using random sequencing."
 RL J. Bacteriol. 175:7918-7930(1993).
 CC -1- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHERIN PROTEINS IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY SIMILARITY).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; U39701; AAC71437.1;
 CC EMBL; U02165; AAD12447.1;
 CC TIGR; MG218;
 KW Cytadherence; Structural protein; Coiled coil; Complete proteome.

FT DOMAIN 28 838 COILED COIL (POTENTIAL).
 FT DOMAIN 914 1591 COILED COIL (POTENTIAL).
 FT DOMAIN 1632 1723 COILED COIL (POTENTIAL).
 FT DOMAIN 1777 1804 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1805 AA; 216252 MW; 11D093AF173284FD CRC64;
 Query Match 5.7%; Score 130; DB 1; Length 1805;
 Best Local Similarity 21.2%; Pred. No. 17;
 Matches 84; Conservative 66; Mismatches 124; Indels 122; Gaps 20;
 Qy 25 KFERQHMDSPDLGTDGDDKANADIGSTEGRTMAISVTMDNILSGFENEYD--VIYLLKPL 82
 Db 1191 KFEKONFD-----TEKQQLVAIKTQCEKLSDEKALKNOKLVELKNL 1232
 Qy 83 AGVYRSLLKQTE-----KNIFTNLNL---NDILNSRLAKRKRYF--LDVLESOL- 126
 Db 1233 SQTILANKNAEYSSQOQLOQKYTNLLDLKENLERTKQOLDK--KRSIFARLTFANDLR 1290
 Qy 127 MQFHSSNEYIIEDSPKLNSEKNTLLSKYKIKESVENDIKFAQBGISYVEK----- 181
 Db 1291 FEKKOLLKAQRIVDDKNRLKENERN-----LHFLSNETERKRAVLEDOISYFEKQKOA 1345
 Qy 182 ---VLAKYKO-----DLESIKKVIKEKEKFPSPPTTPPSAKTDEOKKE----- 224
 Db 1346 TDAIASHKEVKKKEGELQKLLVELETRKTKLNN-----DFAKFSQREEFENQRLKL 1398
 Qy 225 ---SKFLPFTNITNLYNNLVNKIDDDYL-----INLKAINDCN-----V 261
 Db 1399 LELOKTLTOTNSNFKTKAQEIENSYKRGMEELNQKKEFDKNSRLYEFPRKMRDEI 1458
 Qy 262 EKDEAHVKIT-----KLSDLKAIDDKIDLFKNPYDF-----EAIKKLIN---DQTKK-- 305
 Db 1459 ERKESQVKLVLETKQRKANLLEAQANKLNIEKNTIDFEKELKAFKQKVDQIDISTNKQR 1518
 Qy 306 ---DML--GKLLSTGLVQ-----NFPNTIISKLIE 330
 Db 1519 KELNELLNENKLLQOOLIERAINSKDSLUNKKIE 1554

RESULT 36
 GIDA_THEME STANDARD; PRT; 629 AA.
 AC O9KYAL;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Glucose inhibited division protein A.
 GN GIDA OR TM0263.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogae (class); Thermotogales;
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MSE8 / DSM 3109;
 RX MEDLINE-99287316; PubMed-10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RA "Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima."
 RL Nature 399:323-329(1999).
 CC -1- FUNCTION: NOT KNOWN.
 CC -1- SIMILARITY: BELONGS TO THE GIDA FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; U39701; AAC71437.1;
 CC EMBL; U02165; AAD12447.1;
 CC TIGR; MG218;
 KW Cytadherence; Structural protein; Coiled coil; Complete proteome.

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM:protein - protein search, using sw model

Run On: May 12, 2003, 10:15:19 ; Search time 88 Seconds
(without alignments)
1009.163 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 2275

Sequence: 1 MHHHHHHSSGLVPRGSMKE.....TCECTKPDYPLFDGIFCSS 431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:.*
1: sp_archaea:.*
2: sp_bacteria:.*
3: sp_fungi:.*
4: sp_human:.*
5: sp_invertebrate:.*
6: sp_mammal:.*
7: sp_mhc:.*
8: sp_organelle:.*
9: sp_phage:.*
10: sp_plant:.*
11: sp_rodent:.*
12: sp_virus:.*
13: sp_vertebrate:.*
14: sp_unclassified:.*
15: sp_rvirus:.*
16: sp_bacteriap:.*
17: sp_archaeap:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1969.5	86.6	1720	5	Q25922	Q25922 plasmodium
2	1965.5	86.4	652	5	Q25923	Q25923 plasmodium
3	1946.5	85.6	570	5	Q9TYG2	Q9TYG2 plasmodium
4	1946.5	85.6	570	5	Q25968	Q25968 plasmodium
5	1942	85.4	373	5	Q25723	Q25723 plasmodium
6	1937	85.1	569	5	Q25978	Q25978 plasmodium
7	1935	85.1	373	5	Q25722	Q25722 plasmodium
8	1933	85.0	373	5	Q25727	Q25727 plasmodium
9	1931	84.9	569	5	Q25970	Q25970 plasmodium
10	1931	84.9	569	5	Q25980	Q25980 plasmodium
11	1931	84.9	569	5	Q25982	Q25982 plasmodium
12	1929	84.8	373	5	Q43996	Q43996 plasmodium
13	1928	84.7	373	5	Q25728	Q25728 plasmodium
14	1924	84.6	373	5	Q25724	Q25724 plasmodium
15	1922.5	84.5	372	5	Q25717	Q25717 plasmodium
16	1921	84.4	569	5	Q25983	Q25983 plasmodium

17	1919	84.4	373	5	Q43995	Q43995 plasmodium
18	1919	84.4	373	5	Q25721	Q25721 plasmodium
19	1917.5	84.3	372	5	Q25718	Q25718 plasmodium
20	1917.5	84.3	372	5	Q25719	Q25719 plasmodium
21	1917.5	84.3	372	5	Q25720	Q25720 plasmodium
22	1917	84.3	651	5	Q25924	Q25924 plasmodium
23	1915	84.2	569	5	Q25969	Q25969 plasmodium
24	1915	84.2	569	5	Q25974	Q25974 plasmodium
25	1915	84.2	569	5	Q25975	Q25975 plasmodium
26	1915	84.2	569	5	Q25977	Q25977 plasmodium
27	1915	84.2	569	5	Q25979	Q25979 plasmodium
28	1915	84.2	1694	5	Q9NHX1	Q9NHX1 plasmodium
29	1915	84.2	1694	5	Q9TZT5	Q9TZT5 plasmodium
30	1915	84.2	1704	5	Q9TZT4	Q9TZT4 plasmodium
31	1906.5	83.8	372	5	Q25725	Q25725 plasmodium
32	1901.5	83.6	372	5	Q25726	Q25726 plasmodium
33	1900.5	83.5	372	5	Q43997	Q43997 plasmodium
34	1900	83.5	569	5	Q25967	Q25967 plasmodium
35	1178	51.8	539	5	Q25972	Q25972 plasmodium
36	1178	51.8	539	5	Q25981	Q25981 plasmodium
37	1175	51.6	400	5	Q03999	Q03999 plasmodium
38	1175	51.6	539	5	Q25984	Q25984 plasmodium
39	1175	51.6	539	5	Q25966	Q25966 plasmodium
40	1175	51.6	539	5	Q9TYG1	Q9TYG1 plasmodium
41	1175	51.6	539	5	Q25976	Q25976 plasmodium
42	1169	51.4	539	5	Q25971	Q25971 plasmodium
43	1169	51.4	539	5	Q25973	Q25973 plasmodium
44	1165.5	51.2	376	5	Q9BMG8	Q9BMG8 plasmodium
45	1049	46.1	218	5	Q9TVG8	Q9TVG8 plasmodium

ALIGNMENTS

RESULT 1

Q25922 ID Q25922 PRELIMINARY; PRT; 1720-AA.
AC Q25922;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Precursor of the major merozoite surface antigens.
OS Plasmodium falciparum (isolate NF54).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5843;
RN [1]
RP SEQUENCE OF 1069-1720 FROM N.A.
RC STRAIN=NF54;
RA Tolle R., Bujard H., Cooper J.A.;
RL Exp. Parasitol. 0:0-0(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NF54;
RA Tolle R.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC MEDLINE=96123395; PubMed=8577332;
RA Pan W., Tolle R., Bujard H.;
RT "A direct and rapid sequencing strategy for the Plasmodium falciparum antigen gene gp190/MSA1.";
RL Mol. Biochem. Parasitol. 73:241-244(1995).
DR EMBL; 235327; CA84556.1; -;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW Merozoite; Signal.
FT SIGNAL 1 19 POTENTIAL.
SQ SEQUENCE 1720 AA; 195725 MW; 717B2FD1E637A8A3 CRC64;

Query Match 86.6%; Score 1969.5; DB 5; Length 1720;
Best Local Similarity 95.7%; Pred. No. 8.7e-88;
Matches 377; Conservative 7; Mismatches 5; Indels 5; Gaps 1;

```
QY 38 TDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQKQIEKNI 97
Db 1312.SEDNDYLDQVVTCG-----AISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQKQIEKNI 1366
QY 98 FTFNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIESFKLLNSEQKNTLLKS 157
Db 1367 FTFNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIESFKLLNSEQKNTLLKS 1426
QY 158 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAK 217
Db 1427 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAK 1486
QY 218 TDEQKESKFLPFTLTNIETLYNNLVNKIDYLYNLKAKINDCNVEKDEAHVKITKLSDLK 277
Db 1487 TDEQKESKFLPFTLTNIETLYNNLVNKIDYLYNLKAKINDCNVEKDEAHVKITKLSDLK 1546
QY 278 AIDDKIDLFKNPYDFAEIKKLLNDTKKMLGKLLSTGLVQNFNTIISKLEIEGKFQDML 337
Db 1547 AIDDKIDLFKNPYDFAEIKKLLNDTKKMLGKLLSTGLVQNFNTIISKLEIEGKFQDML 1606
QY 338 NISOHCVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPPTCNENNGCCDADA 397
Db 1607 NISOHCVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPPTCNENNGCCDADA 1666
QY 398 TCTEEDSGSSRRKKTCTCKPDSYPLDFGIFCSS 431'
Db 1667 TCTEEDSGSSRRKKTCTCKPDSYPLDFGIFCSS 1700
```

RESULT 2

```
Q25923 ID Q25923 PRELIMINARY; PRT; 652 AA.
AC Q25923;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Merozoite surface antigen 1 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE OF 1-298 FROM N.A.
RC STRAIN=FCH5/NF7;
RX MEDLINE=88142999; PubMed=2449612;
RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
RT Brown G.V., Anders R.F., Kemp D.J.;
RT "Variation in the precursor to the major merozoite surface antigens of
RT Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 27:291-302(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FCH5/NF7;
RX MEDLINE=95354793; PubMed=7628566;
RA Tolle R.;
RT "Plasmodium falciparum: variations within the C-terminal region of
RT merozoite surface antigen-1.";
RL Exp. Parasitol. 81:47-54(1995).
DR EMBL; Z35328; CAA84557.1; -
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1
SQ SEQUENCE 652 AA; 74292 MW; 286A87737B490A62 CRC64;

Query Match 86.4%; Score 1965.5; DB 5; Length 652;
Best Local Similarity 95.4%; Pred. No. 4.9e-88;
Matches 376; Conservative 8; Mismatches 5; Indels 5; Gaps 1;
```

```
QY 38 TDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQKQIEKNI 97
Db 244 SEDNDYLDQVVTCG-----AISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQKQIEKNI 298
QY 98 FTFNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIESFKLLNSEQKNTLLKS 157
Db 299 FTFNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIESFKLLNSEQKNTLLKS 358
QY 158 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAK 217
Db 359 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAK 418
QY 218 TDEQKESKFLPFTLTNIETLYNNLVNKIDYLYNLKAKINDCNVEKDEAHVKITKLSDLK 277
Db 419 TDEQKESKFLPFTLTNIETLYNNLVNKIDYLYNLKAKINDCNVEKDEAHVKITKLSDLK 478
QY 278 AIDDKIDLFKNPYDFAEIKKLLNDTKKMLGKLLSTGLVQNFNTIISKLEIEGKFQDML 337
Db 479 AIDDKIDLFKNPYDFAEIKKLLNDTKKMLGKLLSTGLVQNFNTIISKLEIEGKFQDML 538
QY 338 NISOHCVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPPTCNENNGCCDADA 397
Db 539 NISOHCVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPPTCNENNGCCDADA 598
QY 398 TCTEEDSGSSRRKKTCTCKPDSYPLDFGIFCSS 431
Db 599 TCTEEDSGSSRRKKTCTCKPDSYPLDFGIFCSS 632

RESULT 3
Q9TYG2 ID Q9TYG2 PRELIMINARY; PRT; 570 AA.
AC Q9TYG2;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88011243; PubMed=3079521;
RA Tanabe K., Mackay M., Goman M., Scalfe J.G.;
RT "Allelic Dimorphism in a Surface Antigen Gene of the Malaria Parasite
RT Plasmodium falciparum.";
RL J. Mol. Biol. 195:273-287(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13343; BAA02604.1; -
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 570 AA; 64630 MW; 8674DEC89B2D662A CRC64;

Query Match 85.6%; Score 1946.5; DB 5; Length 570;
Best Local Similarity 94.9%; Pred. No. 3.6e-87;
Matches 374; Conservative 7; Mismatches 8; Indels 5; Gaps 1;
```

```
QY 38 TDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQKQIEKNI 97
Db 162 SEDNDYLDQVVTCG-----AISVTMDNILSGFENEYDVYILKPLAGYVRSLLKQKQIEKNI 216
QY 98 FTFNLNDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIESFKLLNSEQKNTLLKS 157
```

```
Db 217 TTFNLNLNDILNSRLKKRYFLDVLESMLQFKHSSNEYIIEDSFKLNSQKNTLLKS 276
Qy 158 YKIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEEKKFPSSPTTPPSPAK 217
Db 277 YKIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEEKKFPSSPTTPPSPAK 336
Qy 218 TDEQKESKFLPFLTNIETLYNNLNVKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLK 277
Db 337 TDEQKESKFLPFLTNIETLYNNLNVKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLK 396
Qy 278 ATDDKIDLFKNPYDPEATKLLNDOTKDDMLGKLLSTGLVQNFPTIISKLEGGKFDQML 337
Db 397 ATDDKIDLFKNPYDPEATKLLNDOTKDDMLGKLLSTGLVQNFPTIISKLEGGKFDQML 456
Qy 338 NISQHCYKQKCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADA 397
Db 457 NISQHCYKQKCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADA 516
Qy 398 TCTEEDSGSSRRKKTCECTKPDSPYPLFDGIFCSS 431
Db 517 TCTEEDSGSSRRKKTCECTKPDSPYPLFDGIFCSS 550

RESULT 4
Q25968
ID Q25968 PRELIMINARY; PRT; 570 AA.
AC Q25968:
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93293445; PubMed=8515786;
RA Jongwatwises S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates";
RL MOL. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13345; BAA02606.1; -
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 570 AA; 64632 MW; 424BF553CCC2F2BE CRC64;

Query Match 85.6%; Score 1946.5; DB 5; Length 570;
Best Local Similarity 94.9%; Pred. No. 3.6e-87;
Matches 374; Conservative 7; Mismatches 8; Indels 5; Gaps 1;

Qy 38 TDDDDKAMADIGSIEGRGTMAISVTMDNLISGFENYDVYILKPLAGVYRSLLKQIEKNI 97
Db 162 SEDNDEYLDQVVTGE-----AISVTMDNLISGFENYDVYILKPLAGVYRSLLKQIEKNI 216
Qy 98 TTFNLNLNDILNSRLKKRYFLDVLESMLQFKHSSNEYIIEDSFKLNSQKNTLLKS 157
Db 217 TTFNLNLNDILNSRLKKRYFLDVLESMLQFKHSSNEYIIEDSFKLNSQKNTLLKS 276
Qy 158 YKIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEEKKFPSSPTTPPSPAK 217
Db 277 YKIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEEKKFPSSPTTPPSPAK 336
Qy 218 TDEQKESKFLPFLTNIETLYNNLNVKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLK 277
Db 337 TDEQKESKFLPFLTNIETLYNNLNVKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLK 396
Qy 278 ATDDKIDLFKNPYDPEATKLLNDOTKDDMLGKLLSTGLVQNFPTIISKLEGGKFDQML 337
Db 397 ATDDKIDLFKNPYDPEATKLLNDOTKDDMLGKLLSTGLVQNFPTIISKLEGGKFDQML 456
Qy 338 NISQHCYKQKCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADA 397
Db 457 NISQHCYKQKCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADA 516
Qy 398 TCTEEDSGSSRRKKTCECTKPDSPYPLFDGIFCSS 431
Db 517 TCTEEDSGSSRRKKTCECTKPDSPYPLFDGIFCSS 550
```

```
Qy 338 NISQHCYKQKCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADA 397
Db 457 NISQHCYKQKCPENSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDADA 516
Qy 398 TCTEEDSGSSRRKKTCECTKPDSPYPLFDGIFCSS 431
Db 517 TCTEEDSGSSRRKKTCECTKPDSPYPLFDGIFCSS 550

RESULT 5
Q25723
ID Q25723 PRELIMINARY; PRT; 373 AA.
AC Q25723:
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoas M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
RT (MSP-1) in field-derived Plasmodium falciparum";
RL Submitted (FEB-1995) to the EMBL/GenBank/DBSJ databases.
DR EMBL; U20728; AAG62219.1; -
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1
FT NON_TER 373
SQ SEQUENCE 373 AA; 42848 MW; EE9A891631DE174F CRC64;

Query Match 85.4%; Score 1942; DB 5; Length 373;
Best Local Similarity 98.9%; Pred. No. 3.8e-87;
Matches 369; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 58 AISVTMDNLISGFENYDVYILKPLAGVYRSLLKQIEKNIETFNLNLNDILNSRLKKRY 117
Db 1 AISVTMDNLISGFENYDVYILKPLAGVYRSLLKQIEKNIETFNLNLNDILNSRLKKRY 60
Qy 118 FLDVLESMLQFKHSSNEYIIEDSFKLNSQKNTLLKSYKIKESVENDIKFAQEGIS 177
Db 61 FLDVLESMLQFKHSSNEYIIEDSFKLNSQKNTLLKSYKIKESVENDIKFAQEGIS 120
Qy 178 YYEKVLAKYKDDLESIKKVIKEEKKFPSSPTTPPSPAKTDEQKESKFLPFLTNIETL 237
Db 121 YYEKVLAKYKDDLESIKKVIKEEKKFPSSPTTPPSPAKTDEQKESKFLPFLTNIETL 180
Qy 238 YNNLNVKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKATDDKIDLFKNPYDFAIKK 297
Db 181 YNNLNVKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKATDDKIDLFKNPYDFAIKK 240
Qy 298 LINDDTKKDMLGKLLSTGLVQNFPTIISKLEGGKFDQMLNISOHCYKQKCPENSGCFR 357
Db 241 LINDDTKKDMLGKLLSTGLVQNFPTIISKLEGGKFDQMLNISOHCYKQKCPENSGCFR 300
Qy 358 HLDRECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRRKKTCECTK 417
Db 301 HLDRECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSGKKITCECTK 360
Qy 418 PDSPYPLFDGIFCS 430
Db 361 PDSPYPLFDGIFCS 373

RESULT 6
Q25978
ID Q25978 PRELIMINARY; PRT; 569 AA.
AC Q25978;
```



```
Db 1 A1SVTMDNLNLSGFENEDYVYIYKPLAGVYRSKKQIEKNIITFNLDNLNLSRLKRRY 60
118 FLDVLESOLMQFKHSSNEVYIEDSKLNSSEKNTLLSKYKIKESVENDIKFAOEGIS 177
61 FLDVLESOLMQFKHSSNEVYIEDSKLNSSEKNTLLSKYKIKESVENDIKFAOEGIS 120
QY 178 YEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAKTDQKKESKFLPFLTNIETL 237
Db 121 YEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAKTDQKKESKFLPFLTNIETL 180
QY 238 YNNLVNKIDDYLLNLKAKINDCNVDEAHVKITKLSDLKAIDDKIDLKFNKYDFEAIKK 297
Db 181 YNNLVNKIDDYLLNLKAKINDCNVDEAHVKITKLSDLKAIDDKIDLKFNKYDFEAIKK 240
QY 298 LINDDTFKDMLGKLLSTGLVQNPNTIISKLTGKFDMLNISOHOCVKQCPENSGCFR 357
Db 241 LINDDTFKDMLGKLLSTGLVQNPNTIISKLTGKFDMLNISOHOCVKQCPENSGCFR 300
QY 358 HLDERECKLLNYKQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSSRKKITCECTK 417
Db 301 HLDERECKLLNYKQEGDKCVENPNPTCNENNGCGDADATCTEEDSGSSRKKITCECTK 360
QY 418 PDSYPLFDGIFCS 430
Db 361 PDSYPLFDGIFCS 373

RESULT 9
Q25970
ID Q25970 PRELIMINARY; PRT; 569 AA.
AC Q25970;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NAR-2002 (TReMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPI) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; DI3347; BAA02608.1; -.
DR InterPro; IPR000361; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64502 MW; 4D761FF472753142 CRC64;

Query Match 84.9%; Score 1931; DB 5; Length 569;
Best Local Similarity 94.7%; Pred. No. 2e-86;
Matches 373; Conservative 7; Mismatches 8; Indels 6; Gaps 2;

QY 38 TDDDDKAMADIGSIEGRGTMAISVTMDNLNLSGFENEDYVYIYKPLAGVYRSKKQIEKNI 97
Db 162 SEDNDYLDQVVTGE-----A1SVTMDNLNLSGFENEDYVYIYKPLAGVYRSKKQIEKNI 216

QY 98 FTFNLNLDLNSRLKRRKYFDLVLESOLMQFKHSSNEVYIEDSKLNSSEKNTLLKS 157
Db 217 IFTNLNDLILNSRLKRRKYFDLVLESOLMQFKHSSNEVYIEDSKLNSSEKNTLLKS 276

QY 158 YKIKESVENDIKFAOEGISYIEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 217
Db 277 YKIKESVENDIKFAOEGISYIEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 335

QY 218 TDQKKESKFLPFLTNIETLYNNLVNKIDDYLLNLKAKINDCNVDEAHVKITKLSDLK 277
Db 336 TDEQKESKFLPFLTNIETLYNNLVNKIDDYLLNLKAKINDCNVDEAHVKITKLSDLK 395

QY 278 AIDDKIDLKFNKYDFEAIKKLINDDTFKDMLGKLLSTGLVQNPNTIISKLTGKFDML 337
Db 396 AIDDKIDLKFNKYDFEAIKKLINDDTFKDMLGKLLSTGLVQNPNTIISKLTGKFDML 455

QY 338 NISQHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCGDADA 397
Db 456 NISQHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCGDADA 515

QY 398 TCTEEDSGSSRKKITCECTKPDSPYPLFDGIFCSS 431
Db 516 TCTEEDSGSSRKKITCECTKPDSPYPLFDGIFCSS 549
```

```
Db 336 TDEQKESKFLPFLTNIETLYNNLVNKIDDYLLNLKAKINDCNVDEAHVKITKLSDLK 395
QY 278 AIDDKIDLKFNKYDFEAIKKLINDDTFKDMLGKLLSTGLVQNPNTIISKLTGKFDML 337
Db 396 AIDDKIDLKFNKYDFEAIKKLINDDTFKDMLGKLLSTGLVQNPNTIISKLTGKFDML 455
QY 338 NISQHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCGDADA 397
Db 456 NISQHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCGDADA 515
QY 398 TCTEEDSGSSRKKITCECTKPDSPYPLFDGIFCSS 431
Db 516 TCTEEDSGSSRKKITCECTKPDSPYPLFDGIFCSS 549

RESULT 10
Q25980
ID Q25980 PRELIMINARY; PRT; 569 AA.
AC Q25980;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NAR-2002 (TReMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPI) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; DI3353; BAA02614.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64536 MW; 8008861DECECD8DC CRC64;

Query Match 84.9%; Score 1931; DB 5; Length 569;
Best Local Similarity 94.7%; Pred. No. 2e-86;
Matches 373; Conservative 7; Mismatches 8; Indels 6; Gaps 2;

QY 38 TDDDDKAMADIGSIEGRGTMAISVTMDNLNLSGFENEDYVYIYKPLAGVYRSKKQIEKNI 97
Db 162 SEDNDYLDQVVTGE-----A1SVTMDNLNLSGFENEDYVYIYKPLAGVYRSKKQIEKNI 216

QY 98 FTFNLNLDLNSRLKRRKYFDLVLESOLMQFKHSSNEVYIEDSKLNSSEKNTLLKS 157
Db 217 IFTNLNDLILNSRLKRRKYFDLVLESOLMQFKHSSNEVYIEDSKLNSSEKNTLLKS 276

QY 158 YKIKESVENDIKFAOEGISYIEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 217
Db 277 YKIKESVENDIKFAOEGISYIEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 335

QY 218 TDQKKESKFLPFLTNIETLYNNLVNKIDDYLLNLKAKINDCNVDEAHVKITKLSDLK 277
Db 336 TDEQKESKFLPFLTNIETLYNNLVNKIDDYLLNLKAKINDCNVDEAHVKITKLSDLK 395

QY 278 AIDDKIDLKFNKYDFEAIKKLINDDTFKDMLGKLLSTGLVQNPNTIISKLTGKFDML 337
Db 396 AIDDKIDLKFNKYDFEAIKKLINDDTFKDMLGKLLSTGLVQNPNTIISKLTGKFDML 455

QY 338 NISQHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCGDADA 397
Db 456 NISQHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGCGDADA 515

QY 398 TCTEEDSGSSRKKITCECTKPDSPYPLFDGIFCSS 431
Db 516 TCTEEDSGSSRKKITCECTKPDSPYPLFDGIFCSS 549
```

```

RESULT 11
Q25982 ID Q25982 PRELIMINARY; PRT; 569 AA..
AC Q25982;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13354; BAA02615.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64530 MW; 4D734580516D2142 CRC64;

Query Match 84.98; Score 1931; DB 5; Length 569;
Best Local Similarity 94.78; Pred. No. 2e-86;
Matches 373; Conservative 7; Mismatches 8; Indels 6; Gaps 2;

Qy 38 TDDDKADIGSGEGRTMAISVTMDNILSGFENEYDVIYKPLAGYVRSILKKQIEKNI 97
Db 162 SEDNDYLDQVTCGE-----AISVTMDNILSGFENEYDVIYKPLAGYVRSILKKQIEKNI 216
: : : : :
Qy 98 FTFNILNDILNSRLKKRYFDLVLESDFMFKHISSEYIIEDSFKLLNSEQNTLLKS 157
Db 217 ITFNILNDILNSRLKKRYFDLVLESDFMFKHISSEYIIEDSFKLLNSEQNTLLKS 276
: : : : :
Qy 158 YKYIKESVENDIKFAQEGISYKYLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAK 217
Db 277 YKYIKESVENDIKFAQEGISYKYLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAK 335
: : : : :
Qy 218 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYILNLKAKINDCNVEKDEAHVKITKLSDLK 277
Db 336 TDEQKESKFLPFLTNIETLYNNLVNKIDDDYILNLKAKINDCNVEKDEAHVKITKLSDLK 395
: : : : :
Qy 278 AIDDKIDLFKNPYDFEAIKKLINDDTKMDLGLKLLSTGLVQNFNTIISKIEGKFDQML 337
Db 396 AIDDKIDLFKNPYDFEAIKKLINDDTKMDLGLKLLSTGLVQNFNTIISKIEGKFDQML 455
: : : : :
Qy 338 NISQHCYKVKCPCPNSGCFRHLDERECKCLLNKQEGDKCVENPPTCNENNGCCDADA 397
Db 456 NISQHCYKVKCPCPNSGCFRHLDERECKCLLNKQEGDKCVENPPTCNENNGCCDADA 515
: : : : :
Qy 398 TCTEDSGSSRRKKTCTCKPDSYPLFDGIFCSS 431
Db 516 TCTEDSGSSRRKKTCTCKPDSYPLFDGIFCSS 549
: : : : :

RESULT 12
Q43996 ID Q43996 PRELIMINARY; PRT; 373 AA..
AC Q43996;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein-1 (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

```

```

OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=KENYA-2;
RX MEDLINE=98319411; PubMed=9657329;
RA Qari S.H., Shi Y.P., Goldman I.F., Nahlen B.L., Tibayrenc M.,
RA Lal A.A.;
RT "Predicted and observed alleles of Plasmodium falciparum merozoite
RT surface protein-1 (MSP-1), a potential malaria vaccine antigen."
RL Mol. Biochem. Parasitol. 92:241-252(1998).
DR EMBL: AF040568; AAC39098.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
FT NON_TER 1
FT NON_TER 373
SQ SEQUENCE 373 AA; 42902 MW; 86CD4B721E605A5F CRC64;

Query Match 84.88; Score 1929; DB 5; Length 373;
Best Local Similarity 98.9%; Pred. No. 1.6e-86;
Matches 369; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 58 AISVTMDNILSGFENEYDVIYKPLAGYVRSILKKQIEKNIETFNILNDILNSRLKKRY 117
Db 1 AISVTMDNILSGFENEYDVIYKPLAGYVRSILKKQIEKNIETFNILNDILNSRLKKRY 60
: : : : :
Qy 118 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 177
Db 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 120
: : : : :
Qy 178 YKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 237
Db 121 YKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180
: : : : :
Qy 238 YNNLVNKIDDDYILNLKAKINDCNVEKDEAHVKITKLSDLKAIIDKIDLFKNPYDEAIKK 297
Db 181 YNNLVNKIDDDYILNLKAKINDCNVEKDEAHVKITKLSDLKAIIDKIDLFKNPYDEAIKK 240
: : : : :
Qy 298 LINDDTKMDLGLKLLSTGLVQNFNTIISKIEGKFDQMLNISQHCYKVKCPCPNSGCFR 357
Db 241 LINDDTKMDLGLKLLSTGLVQNFNTIISKIEGKFDQMLNISQHCYKVKCPCPNSGCFR 300
: : : : :
Qy 358 HLDRECECKCLLNKQEGDKCVENPPTCNENNGCCDADATCTEDSGSSRRKKTCTCK 417
Db 301 HLDRECECKCLLNKQEGDKCVENPPTCNENNGCCDADATCTEDSGSSRRKKTCTCK 360
: : : : :
Qy 418 PDSYPLFDGIFCS 430
Db 361 SDSYPLFDGIFCS 373
: : : : :

RESULT 13
Q25728 ID Q25728 PRELIMINARY; PRT; 373 AA..
AC Q25728;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Pova M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
RT (MSP-1) in field-derived Plasmodium falciparum."
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U20733; AAA62224.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
FT NON_TER 1
FT NON_TER 373
SQ SEQUENCE 373 AA; 42902 MW; 86CD4B721E605A5F CRC64;

Query Match 84.88; Score 1929; DB 5; Length 373;
Best Local Similarity 98.9%; Pred. No. 1.6e-86;
Matches 369; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 58 AISVTMDNILSGFENEYDVIYKPLAGYVRSILKKQIEKNIETFNILNDILNSRLKKRY 117
Db 1 AISVTMDNILSGFENEYDVIYKPLAGYVRSILKKQIEKNIETFNILNDILNSRLKKRY 60
: : : : :
Qy 118 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 177
Db 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 120
: : : : :
Qy 178 YKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 237
Db 121 YKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180
: : : : :
Qy 238 YNNLVNKIDDDYILNLKAKINDCNVEKDEAHVKITKLSDLKAIIDKIDLFKNPYDEAIKK 297
Db 181 YNNLVNKIDDDYILNLKAKINDCNVEKDEAHVKITKLSDLKAIIDKIDLFKNPYDEAIKK 240
: : : : :
Qy 298 LINDDTKMDLGLKLLSTGLVQNFNTIISKIEGKFDQMLNISQHCYKVKCPCPNSGCFR 357
Db 241 LINDDTKMDLGLKLLSTGLVQNFNTIISKIEGKFDQMLNISQHCYKVKCPCPNSGCFR 300
: : : : :
Qy 358 HLDRECECKCLLNKQEGDKCVENPPTCNENNGCCDADATCTEDSGSSRRKKTCTCK 417
Db 301 HLDRECECKCLLNKQEGDKCVENPPTCNENNGCCDADATCTEDSGSSRRKKTCTCK 360
: : : : :
Qy 418 PDSYPLFDGIFCS 430
Db 361 SDSYPLFDGIFCS 373
: : : : :

```

```
FT NON_TER 373 373
SQ SEQUENCE 373 AA; 42815 MW; 9CF4DDAF38CC4054 CRC64;

Query Match 84.7%; Score 1928; DB 5; Length 373;
Best Local Similarity 98.7%; Pred. No. 1.8e-86;
Matches 368; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 58 A1SVTMDNLTSGFENEYDVIYLLKPLAGVYRSLLKQIEKNIFTNLTNDILNSRLKRRKY 117
Db 1 A1SVTMDNLTSGFENEYDVIYLLKPLAGVYRSLLKQIEKNIFTNLTNDILNSRLKRRKY 60
QY 118 FLDVLES DLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 177
Db 61 FLDVLES DLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 120
QY 178 YVEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTPSPAKTDEQKESKFLPFLTNIETL 237
Db 121 YVEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTPSPAKTDEQKESKFLPFLTNIETL 180
QY 238 YNNLVNKIDDIYLLINKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 297
Db 180 YNNLVNKIDDIYLLINKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 239
QY 298 LINDDTKKDMLGKLLSTGLVQNFPTIISKLIEGKFQDMLNISOHCVKKQCPENSGGCFR 357
Db 240 LINDDTKKDMLGKLLSTGLVQNFPTIISKLIEGKFQDMLNISOHCVKKQCPENSGGCFR 300
QY 358 HLDERECKCLLNKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSSRKKITCECTK 417
Db 301 HLDERECKCLLNKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
QY 418 PDSYPLFDGIFCS 430
Db 361 PDSYPLFDGIFCS 373

RESULT 14
Q25724 PRELIMINARY; PRT; 373 AA.
AC Q25724;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoas M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
RT (MSP-1) in field-derived Plasmodium falciparum.";
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20729; AAA62220.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1 373
FT SEQUENCE 373 AA; 42839 MW; A21A91B2076F415A CRC64;

Query Match 84.6%; Score 1924; DB 5; Length 373;
Best Local Similarity 98.4%; Pred. No. 2.8e-86;
Matches 367; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 58 A1SVTMDNLTSGFENEYDVIYLLKPLAGVYRSLLKQIEKNIFTNLTNDILNSRLKRRKY 117
Db 1 A1SVTMDNLTSGFENEYDVIYLLKPLAGVYRSLLKQIEKNIFTNLTNDILNSRLKRRKY 60
QY 118 FLDVLES DLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 177
Db 61 FLDVLES DLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 120
```

```
QY 178 YVEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTPSPAKTDEQKESKFLPFLTNIETL 237
Db 121 YVEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTPSPAKTDEQKESKFLPFLTNIETL 180
QY 238 YNNLVNKIDDIYLLINKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 297
Db 181 YNNLVNKIDDIYLLINKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
QY 298 LINDDTKKDMLGKLLSTGLVQNFPTIISKLIEGKFQDMLNISOHCVKKQCPENSGGCFR 357
Db 241 LINDDTKKDMLGKLLSTGLVQNFPTIISKLIEGKFQDMLNISOHCVKKQCPENSGGCFR 300
QY 358 HLDERECKCLLNKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSSRKKITCECTK 417
Db 301 HLDERECKCLLNKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
QY 418 PDSYPLFDGIFCS 430
Db 361 PDSYPLFDGIFCS 373

RESULT 15
Q25717 PRELIMINARY; PRT; 372 AA.
AC Q25717;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoas M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
RT (MSP-1) in field-derived Plasmodium falciparum.";
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20653; AAA62213.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1 372
FT SEQUENCE 372 AA; 42687 MW; 3F2BF1152598FB10 CRC64;

Query Match 84.5%; Score 1922.5; DB 5; Length 372;
Best Local Similarity 98.9%; Pred. No. 3.3e-86;
Matches 369; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 58 A1SVTMDNLTSGFENEYDVIYLLKPLAGVYRSLLKQIEKNIFTNLTNDILNSRLKRRKY 117
Db 1 A1SVTMDNLTSGFENEYDVIYLLKPLAGVYRSLLKQIEKNIFTNLTNDILNSRLKRRKY 60
QY 118 FLDVLES DLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 177
Db 61 FLDVLES DLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 120
QY 178 YVEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTPSPAKTDEQKESKFLPFLTNIETL 237
Db 121 YVEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTPSPAKTDEQKESKFLPFLTNIETL 179
QY 238 YNNLVNKIDDIYLLINKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 297
Db 180 YNNLVNKIDDIYLLINKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 239
QY 298 LINDDTKKDMLGKLLSTGLVQNFPTIISKLIEGKFQDMLNISOHCVKKQCPENSGGCFR 357
Db 240 LINDDTKKDMLGKLLSTGLVQNFPTIISKLIEGKFQDMLNISOHCVKKQCPENSGGCFR 300
QY 358 HLDERECKCLLNKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSSRKKITCECTK 417
```

```
Db 300 HLDRECKLLNYKQEGDKCVENPNTCNENGGCCADATCTEEDSGSRKKITCCTK 359
Qy 418 PDSYPLFDGIFCS 430
Db 360 PDSYPLFDGIFCS 372

RESULT 16
Q25983 PRELIMINARY; PRT; 569 AA.
AC Q25983;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwatitwies S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13355; BAA02616.1;
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
FT SEQUENCE 569 AA; 64491 MW; 57AGB62FF72CE885 CRC64;

Query Match 84.4%; Score 1921; DB 5; Length 569;
Best Local Similarity 94.2%; Pred. No. 6.1e-86;
Matches 371; Conservative 8; Mismatches 9; Indels 6; Gaps 2;

Qy 38 TDDDKAMADIGTEGRGTMAISVTMDNLSGFENEDVYILKPLAGVYRSLLKQIEKNI 97
Db 162 SEDNEDVDQVTGE-----AISTMDNLSGFENEDVYILKPLAGVYRSLLKQIEKNI 216
Qy 98 FTFNLNNDILNSRLKRRKRYFLDVLSDLMQFKHISSENYIIEDSFKLLNSEQKNTLLKS 157
Db 217 FTFNLNNDILNSRLKRRKRYFLDVLSDLMQFKHISSENYIIEDSFKLLNSEQKNTLLKS 276
Qy 158 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAK 217
Db 277 YKIKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAK 335
Qy 218 TDEQKESKFLPFLTNTETLYNVLNKKIDDDYLNLKAKINDCNVKEKDEAHVKITKLSDLK 277
Db 336 TDEQKESKFLPFLTNTETLYNVLNKKIDDDYLNLKAKINDCNVKEKDEAHVKITKLSDLK 395
Qy 278 AIDDKIDLFKNPYDFAIKKLNDDTKKMLGKLLSTGLVQNFNTIISKIEGKFQDML 337
Db 396 AIDDKIDLFKNPYDFAIKKLNDDTKKMLGKLLSTGLVQNFNTIISKIEGKFQDML 455
Qy 338 NISQHCVKKQCPNSGCCFRLHLDRECKCLLYNKQEGDKCVENPNTCNENGGCCADA 397
Db 456 NISQHCVKKQCPNSGCCFRLHLDRECKCLLYNKQEGDKCVENPNTCNENGGCCADA 515
Qy 398 TCTEEDSGSRKKITCCTKPDYPLFDGIFCSS 431
Db 516 KCTEEDSGSGNGKKTCTCFDPSYPLFDGIFCSS 549

RESULT 17
Q43995 PRELIMINARY; PRT; 373 AA.
AC Q43995;
DT 01-JUN-1998 (Tremblrel. 06, Created)
```

```
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DE Merozoite surface protein-1 (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=KENYA-1;
RX MEDLINE=98319411; PubMed=9657329;
RA Qari S.H., Shi Y.P., Goldman I.F., Nahlen B.L., Tibayrenc M.,
RA Lal A.A.;
RT "Predicted and observed alleles of Plasmodium falciparum merozoite
RT surface protein-1 (MSP-1), a potential malaria vaccine antigen."
RL Mol. Biochem. Parasitol. 92:241-252(1998).
DR EMBL: AF040567; AAC39097.1;
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
FT NON_TER 1
FT SEQUENCE 373 AA; 42873 MW; 886CF169A7AF5194 CRC64;

Query Match 84.4%; Score 1919; DB 5; Length 373;
Best Local Similarity 98.7%; Pred. No. 4.9e-86;
Matches 368; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 58 AISTMDNLSGFENEDVYILKPLAGVYRSLLKQIEKNIETFNLDNLNLDILNSRLKRRKY 117
Db 1 AISTMDNLSGFENEDVYILKPLAGVYRSLLKQIEKNIETFNLDNLNLDILNSRLKRRKY 60
Qy 118 FLVDLESDLMQFKHISSENYIIEDSFKLLNSEQKNTLLSKYKIKESVENDIKFAQEGIS 177
Db 61 FLVDLESDLMQFKHISSENYIIEDSFKLLNSEQKNTLLSKYKIKESVENDIKFAQEGIS 120
Qy 178 YKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 237
Db 121 YKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180
Qy 238 YNLNKKIDDDYLNLKAKINDCNVKEKDEAHVKITKLSDLKAIIDDKIDLFKNPYDFAIKK 297
Db 181 YNLNKKIDDDYLNLKAKINDCNVKEKDEAHVKITKLSDLKAIIDDKIDLFKNPYDFAIKK 240
Qy 298 LINDTKKMLGKLLSTGLVQNFNTIISKIEGKFQDMLNLSQHCVKKQCPNSGCCFR 357
Db 241 LINDTKKMLGKLLSTGLVQNFNTIISKIEGKFQDMLNLSQHCVKKQCPNSGCCFR 300
Qy 358 HLDRECKCLLYNKQEGDKCVENPNTCNENGGCCADATCTEEDSGSRKKITCCTK 417
Db 301 HLDRECKCLLYNKQEGDKCVENPNTCNENGGCCADATCTEEDSGSRKKITCCTK 360
Qy 418 PDSYPLFDGIFCS 430
Db 361 PDSYPLFDGIFCS 373

RESULT 18
Q25721 PRELIMINARY; PRT; 373 AA.
AC Q25721;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RA "Sequence of the C-terminal region of merozoite surface protein-1
```

RT (MSP-1) in field-derived Plasmodium falciparum.";
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U20726; AAA62217.1; -;
 DR InterPro: IPR000561; EGF-like.
 DR Pfam: PF00008; EGF; 1;
 FT NON_TER 1 373
 SQ SEQUENCE 373 AA; 42848 MW; EE0700233D7F4D4E CRC64;

Query Match 84.4%; Score 1919; DB 5; Length 373;
 Best Local Similarity 98.1%; Pred. No. 4.9e-86;
 Matches 366; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 58 A1SVTMDN1LGFENEYDVIYKPLAGVYRSLKQIEKNIFTN1LNLDILNSRLKRRY 117
 Db 1 A1SVTMDN1LGFENEYDVIYKPLAGVYRSLKQIEKNIFTN1LNLDILNSRLKRRY 60
 QY 118 FLDVLESOLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAOEGIS 177
 Db 61 FLDVLESOLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAOEGIS 120
 QY 178 YVEKVLAKYKDDLESIKKVIKEEKFPSPPTTPSPAKTDEQKESKFLPLTNIELT 237
 Db 121 YVEKVLAKYKDDLESIKKVIKEEKFPSPPTTPSPAKTDEQKESKFLPLTNIELT 180
 QY 238 YNNLVNKIDYILN1LAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFEAIKK 297
 Db 181 YNNLVNKIDYILN1LAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFEAIKK 240
 QY 298 LINDOTKDKMLGKLLSTGLVQNFPTIISKLEGGFQDMLN1SOHCVKKQCPENSGCPR 357
 Db 241 LINDOTKDKMLGKLLSTGLVQNFPTIISKLEGGFQDMLN1SOHCVKKQCPENSGCPR 300
 QY 358 HLDERECKCLLNYKQEGDKCVENPNTCNENNGCGDADATCTEEDSGSRKKITCECTK 417
 Db 301 HLDERECKCLLNYKQEGDKCVENPNTCNENNGCGDADATCTEEDSGSGNGKVCYCECTK 360
 QY 418 PDSYPLFDGIFCS 430
 Db 361 PDSYPLFDGIFCS 373

RESULT 19

Q25718

ID Q25718 PRELIMINARY; PRT; 372 AA.

AC Q25718;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).

GN MSP-1.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-INFECTED HUMAN BLOOD SAMPLE;

RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;

RT "Sequence of the C-terminal region of merozoite surface protein-1

RT (MSP-1) in field-derived Plasmodium falciparum.";

RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.

DR EMBL; U20654; AAA62214.1; -;

DR InterPro: IPR000561; EGF-like.

DR Pfam: PF00008; EGF; 1.

FT NON_TER 1 372

SQ SEQUENCE 372 AA; 42660 MW; E9F088252598FB0B CRC64;

Query Match 84.3%; Score 1917.5; DB 5; Length 372;

Best Local Similarity 98.7%; Pred. No. 5.8e-86;

Matches 368; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 58 A1SVTMDN1LGFENEYDVIYKPLAGVYRSLKQIEKNIFTN1LNLDILNSRLKRRY 117

Db 1 A1SVTMDN1LGFENEYDVIYKPLAGVYRSLKQIEKNIFTN1LNLDILNSRLKRRY 60
 QY 118 FLDVLESOLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAOEGIS 177
 Db 61 FLDVLESOLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAOEGIS 120
 QY 178 YVEKVLAKYKDDLESIKKVIKEEKFPSPPTTPSPAKTDEQKESKFLPLTNIELT 237
 Db 121 YVEKVLAKYKDDLESIKKVIKEEKFPSPPTTPSPAKTDEQKESKFLPLTNIELT 179
 QY 238 YNNLVNKIDYILN1LAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFEAIKK 297
 Db 180 YNNLVNKIDYILN1LAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFEAIKK 239
 QY 298 LINDOTKDKMLGKLLSTGLVQNFPTIISKLEGGFQDMLN1SOHCVKKQCPENSGCPR 357
 Db 240 LINDOTKDKMLGKLLSTGLVQNFPTIISKLEGGFQDMLN1SOHCVKKQCPENSGCPR 299
 QY 358 HLDERECKCLLNYKQEGDKCVENPNTCNENNGCGDADATCTEEDSGSRKKITCECTK 417
 Db 300 HLDERECKCLLNYKQEGDKCVENPNTCNENNGCGDADATCTEEDSGSGSRKKITCECTK 359
 QY 418 PDSYPLFDGIFCS 430
 Db 360 PDSYPLFDGIFCS 372

RESULT 20

Q25719

ID Q25719 PRELIMINARY; PRT; 372 AA.

AC Q25719;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).

GN MSP-1.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-INFECTED HUMAN BLOOD SAMPLE;

RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;

RT "Sequence of the C-terminal region of merozoite surface protein-1

RT (MSP-1) in field-derived Plasmodium falciparum.";

RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.

DR EMBL; U20655; AAA62215.1; -;

DR InterPro: IPR000561; EGF-like.

DR Pfam: PF00008; EGF; 1.

FT NON_TER 1 372

SQ SEQUENCE 372 AA; 42686 MW; 583A33972DB05FB4 CRC64;

Query Match 84.3%; Score 1917.5; DB 5; Length 372;

Best Local Similarity 98.7%; Pred. No. 5.8e-86;

Matches 368; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 58 A1SVTMDN1LGFENEYDVIYKPLAGVYRSLKQIEKNIFTN1LNLDILNSRLKRRY 117

Db 1 A1SVTMDN1LGFENEYDVIYKPLAGVYRSLKQIEKNIFTN1LNLDILNSRLKRRY 60

QY 118 FLDVLESOLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAOEGIS 177

Db 61 FLDVLESOLMQFKHSSNEYIIEDSFKLLNSEQKNTLLSKYIKESVENDIKFAOEGIS 120

QY 178 YVEKVLAKYKDDLESIKKVIKEEKFPSPPTTPSPAKTDEQKESKFLPLTNIELT 237

Db 121 YVEKVLAKYKDDLESIKKVIKEEKFPSPPTTPSPAKTDEQKESKFLPLTNIELT 179

QY 238 YNNLVNKIDYILN1LAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFEAIKK 297

Db 180 YNNLVNKIDYILN1LAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPKNPYDFEAIKK 239

```
QY 298 LINDTTKMDLGLKSLGVLQVNFNTIISKLIIEGKFQDMLNISQHCQVKKQCPNSGCFR 357
|||||
Db 240 LINDTTKMDLGLKSLGVLQVNFNTIISKLIIEGKFQDMLNISQHCQVKKQCPNSGCFR 299

QY 358 HLDRECKCLLNYKQBGDKCVENPNTCNENNGCGDADATCTEEDSGSRKKITCCECTK 417
|||||
Db 300 HLDRECKCLLNYKQBGDKCVENPNTCNENNGCGDADATCTEEDSGSRKKITCCECTK 359

QY 418 PDSYPLFDGIFCS 430
|||||
Db 360 PDSYPLFDGIFCS 372

RESULT 21
Q25720 PRELIMINARY; PRT; 372 AA.
AC Q25720;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RO-71; PubMed=1592091;
RX MEDLINE=92275047; PubMed=1592091;
RA Olafsson P., Matile H., Certa U.;
RT "Plasmodium falciparum: the repetitive MSA-1 surface protein of the
RT RO-71 isolate is recognized by mouse antibody against the
RT nonrepetitive repeat block of RO-33.";
RL Exp. Parasitol. 74:381-389(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RO-71; PubMed=7628566;
RX MEDLINE=95354793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
RT merozoite surface antigen-1.";
RL Exp. Parasitol. 81:47-54(1995).
DR EMBL: 235329; CAAB4558.1;
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
FT NON_TER 1
FT NON_SEQUENCE 372
SQ SEQUENCE 372 AA; 42686 MW; 583A33972DB05FB4 CRC64;

Query Match 84.3%; Score 1917.5; DB 5; Length 372;
Best Local Similarity 98.7%; Pred. No. 5.8e-86;
Matches 368; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 58 AISVTMDNILSGFENEYDVLYLPLAGVYRSLKKQIEKNITFTNLNDILNSRLKKRY 117
|||||
Db 1 AISVTMDNILSGFENEYDVLYLPLAGVYRSLKKQIEKNITFTNLNDILNSRLKKRY 60

QY 118 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQKNTLLSKYKIKESVENDIKFAQEGIS 177
|||||
Db 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQKNTLLSKYKIKESVENDIKFAQEGIS 120

QY 178 YEKVLAKYKDDLESIKKVIKEEKEFPSSPTTPPSPAKTDEQKESKFLPFLTNIETL 237
|||||
Db 121 YEKVLAKYKDDLESIKKVIKEEKE-FPSSPTTPPSPAKTDEQKESKFLPFLTNIETL 179

QY 238 YNNLVNKIDDYLYNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNPYDFAEIKK 297
|||||
Db 180 YNNLVNKIDDYLYNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNPYDFAEIKK 239

QY 298 LINDTTKMDLGLKSLGVLQVNFNTIISKLIIEGKFQDMLNISQHCQVKKQCPNSGCFR 357
|||||
Db 240 LINDTTKMDLGLKSLGVLQVNFNTIISKLIIEGKFQDMLNISQHCQVKKQCPNSGCFR 299

QY 358 HLDRECKCLLNYKQBGDKCVENPNTCNENNGCGDADATCTEEDSGSRKKITCCECTK 417
|||||
Db 300 HLDRECKCLLNYKQBGDKCVENPNTCNENNGCGDADATCTEEDSGSRKKITCCECTK 359

QY 418 PDSYPLFDGIFCS 430
|||||
Db 360 PDSYPLFDGIFCS 372
```

```
RESULT 22
Q25924 PRELIMINARY; PRT; 651 AA.
AC Q25924;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface antigen 1 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RO-71;
RX MEDLINE=92275047; PubMed=1592091;
RA Olafsson P., Matile H., Certa U.;
RT "Plasmodium falciparum: the repetitive MSA-1 surface protein of the
RT RO-71 isolate is recognized by mouse antibody against the
RT nonrepetitive repeat block of RO-33.";
RL Exp. Parasitol. 74:381-389(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RO-71; PubMed=7628566;
RX MEDLINE=95354793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
RT merozoite surface antigen-1.";
RL Exp. Parasitol. 81:47-54(1995).
DR EMBL: 235329; CAAB4558.1;
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
FT NON_TER 1
FT NON_SEQUENCE 651 AA; 74134 MW; AA2137E699255150 CRC64;

Query Match 84.3%; Score 1917; DB 5; Length 651;
Best Local Similarity 93.9%; Pred. No. 1.1e-85;
Matches 370; Conservative 9; Mismatches 9; Indels 6; Gaps 2;

QY 38 TDDDKAMADIGSTIEGRGTMAISVTMDNILSGFENEYDVLYLPLAGVYRSLKKQIEKNI 97
|||||
Db 244 SEDNDEYLDQVVTGE-----AISVTMDNILSGFENEYDVLYLPLAGVYRSLKKQIEKNI 298

QY 98 FTFNLNDILNSRLKKRYFLDVLESDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKS 157
|||||
Db 299 FTFNLNDILNSRLKKRYFLDVLESDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKS 358

QY 158 YKIKESVENDIKFAQEGISYKVKLAKYKDDLESIKKVIKEEKEFPSSPTTPPSPAK 217
|||||
Db 359 YKIKESVENDIKFAQEGISYKVKLAKYKDDLESIKKVIKEEKE-FPSSPTTPPSPAK 417

QY 218 TDEQKESKFLPFLTNIETLYNLVKNIDDYLYNLKAKINDCNVKEDEAHVKITKLSDLK 277
|||||
Db 418 TDEQKESKFLPFLTNIETLYNLVKNIDDYLYNLKAKINDCNVKEDEAHVKITKLSDLK 477

QY 278 AIDDKIDLFKNPYDFAEIKKIDNDTKOMLGLKSLGVLQVNFNTIISKLIIEGKFQDML 337
|||||
Db 478 AIDDKIDLFKNPYDFAEIKKIDNDTKOMLGLKSLGVLQVNFNTIISKLIIEGKFQDML 537

QY 338 NISOHQCVKKQCPNSGCFRHLDERECKCLLNYKQBGDKCVENPNTCNENNGCGDADA 397
|||||
Db 538 NISOHQCVKKQCPNSGCFRHLDERECKCLLNYKQBGDKCVENPNTCNENNGCGDADA 597

QY 398 TCTEEDSGSRKKITCCECTKPDSPYPLFDGIFCSS 431
|||||
Db 598 KCTEEDSGSGKKITCCECTKPDSPYPLFDGIFCSS 631

RESULT 23
Q25969 PRELIMINARY; PRT; 569 AA.
AC Q25969;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
```

```

DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13346; BAA02607.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64485 MW; A26C45914AF01396 CRC64;

Query Match 84.2%; Score 1915; DB 5; Length 569;
Best Local Similarity 93.9%; Pred. No. 1.2e-85;
Matches 370; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

QY 38 TDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVYILKPLAGVYRSLLKKQIEKNI 97
Db 162 SEDNDEYLDQVVTGE-----AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKKQIEKNI 216
QY 98 FTFNLNLDILNSRLKRRKYFLDVLESMDLMQFKHISSEYIIEDSPKLLNSEQNTLLKS 157
Db 217 ITFNLNLDILNSRLKRRKYFLDVLESMDLMQFKHISSEYIIEDSPKLLNSEQNTLLKS 276
QY 158 YKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEKEKFPSSPPTTPPSPAK 217
Db 277 YKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEKEKFPSSPPTTPPSPAK 335
QY 218 TDQKESKFLPPLTNIETLYNNLVNKIDYLYNLKAKINDCNVKEDEAHVKITKLSDLK 277
Db 336 TDQKESKFLPPLTNIETLYNNLVNKIDYLYNLKAKINDCNVKEDEAHVKITKLSDLK 395
QY 278 AIDDKTDLKPNVDFEAIKKLINDDTFKDMGLKLLSTGLVQNPNTIISKLEGRFQDML 337
Db 396 AIDDKTDLKPNVDFEAIKKLINDDTFKDMGLKLLSTGLVQNPNTIISKLEGRFQDML 455
QY 338 NISQHCVKKQCPENSGCFRHLDERECKCLLYKQEGDKCVENPNPTCNENNGGCDADA 397
Db 456 NISQHCVKKQCPENSGCFRHLDERECKCLLYKQEGDKCVENPNPTCNENNGGCDADA 515
QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431
Db 516 KCTEEDSGSGKKITCECTKPDSPYPLFDGIFCSS 549

RESULT 24
Q25974
ID Q25974 PRELIMINARY; PRT; 569 AA.
AC Q25974;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13346; BAA02607.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64459 MW; ED4284B2867C9703 CRC64; 1

```

```

RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13348; BAA02609.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;

Query Match 84.2%; Score 1915; DB 5; Length 569;
Best Local Similarity 93.9%; Pred. No. 1.2e-85;
Matches 370; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

QY 38 TDDDKAMADIGSIEGRGTMAISVTMDNILSGFENEYDVYILKPLAGVYRSLLKKQIEKNI 97
Db 162 SEDNDEYLDQVVTGE-----AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKKQIEKNI 216
QY 98 FTFNLNLDILNSRLKRRKYFLDVLESMDLMQFKHISSEYIIEDSPKLLNSEQNTLLKS 157
Db 217 ITFNLNLDILNSRLKRRKYFLDVLESMDLMQFKHISSEYIIEDSPKLLNSEQNTLLKS 276
QY 158 YKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEKEKFPSSPPTTPPSPAK 217
Db 277 YKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEKEKFPSSPPTTPPSPAK 335
QY 218 TDQKESKFLPPLTNIETLYNNLVNKIDYLYNLKAKINDCNVKEDEAHVKITKLSDLK 277
Db 336 TDQKESKFLPPLTNIETLYNNLVNKIDYLYNLKAKINDCNVKEDEAHVKITKLSDLK 395
QY 278 AIDDKTDLKPNVDFEAIKKLINDDTFKDMGLKLLSTGLVQNPNTIISKLEGRFQDML 337
Db 396 AIDDKTDLKPNVDFEAIKKLINDDTFKDMGLKLLSTGLVQNPNTIISKLEGRFQDML 455
QY 338 NISQHCVKKQCPENSGCFRHLDERECKCLLYKQEGDKCVENPNPTCNENNGGCDADA 397
Db 456 NISQHCVKKQCPENSGCFRHLDERECKCLLYKQEGDKCVENPNPTCNENNGGCDADA 515
QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431
Db 516 KCTEEDSGSGKKITCECTKPDSPYPLFDGIFCSS 549

RESULT 25
Q25975
ID Q25975 PRELIMINARY; PRT; 569 AA.
AC Q25975;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13349; BAA02610.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64459 MW; ED4284B2867C9703 CRC64; 1

```



```
Db 162 SEDNDEYLDQVVTGE-----AISTVMDNILSGFENEYDVIYKPLAGYVRSLLKKQIEKNI 216
Qy 98 FTNLLNLDILNSRLKRRKYFDVLESQKHHISSNEYIIDSFKLLNSEQKNTLLKS 157
Db 217 ITFNLNLDILNSRLKRRKYFDVLESQKHHISSNEYIIDSFKLLNSEQKNTLLKS 276
Qy 158 YKYIKESVENDIKFAQEGISYEVYKLAQKDDLESIKKVIKEEKEKFPSSPTTTPSPAK 217
Db 277 YKYIKESVENDIKFAQEGISYEVYKLAQKDDLESIKKVIKEEKEKFPSSPTTTPSPAK 335
Qy 218 TDEQKESKFLPFLTNTIETLNNLVNKIDDDYLLNKKAKINDCNVEKDEAHVKITKLSDLK 277
Db 336 TDEQKESKFLPFLTNTIETLNNLVNKIDDDYLLNKKAKINDCNVEKDEAHVKITKLSDLK 395
Qy 278 AIDDKIDLFKNPYDFAEATKLLINDTCKKMLGKLLSTGLVONFPNTIISKLEIEGKFQDML 337
Db 455 AIDDKIDLFKNPYDFAEATKLLINDTCKKMLGKLLSTGLVONFPNTIISKLEIEGKFQDML 455
Qy 338 NISQHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNNGGCCADA 397
Db 456 NISQHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNNGGCCADA 515
Qy 398 TCTEEDSGSSRRKKTCTCTKPDSPYPLFDGIFCSS 431
Db 516 KCTEEDSGSNGKKITCTCTKPDSPYPLFDGIFCSS 549

RESULT 26
Q25977 ID Q25977 PRELIMINARY; PRT; 569 AA.
AC Q25977;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13352; BAA02611.1; -
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64485 MW; A26C45914AF01396 CRC64;

Query Match 84.28; Score 1915; DB 5; Length 569;
Best Local Similarity 93.98; Pred. No. 1.2e-85;
Matches 370; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

Qy 38 TDDDDKAMADIGTSGRGCTMAISVTMDNILSGFENEYDVIYKPLAGYVRSLLKKQIEKNI 97
Db 162 SEDNDEYLDQVVTGE-----AISTVMDNILSGFENEYDVIYKPLAGYVRSLLKKQIEKNI 216
Qy 98 FTNLLNLDILNSRLKRRKYFDVLESQKHHISSNEYIIDSFKLLNSEQKNTLLKS 157
Db 217 ITFNLNLDILNSRLKRRKYFDVLESQKHHISSNEYIIDSFKLLNSEQKNTLLKS 276
Qy 158 YKYIKESVENDIKFAQEGISYEVYKLAQKDDLESIKKVIKEEKEKFPSSPTTTPSPAK 217
Db 277 YKYIKESVENDIKFAQEGISYEVYKLAQKDDLESIKKVIKEEKEKFPSSPTTTPSPAK 335
Qy 218 TDEQKESKFLPFLTNTIETLNNLVNKIDDDYLLNKKAKINDCNVEKDEAHVKITKLSDLK 277
Db 336 TDEQKESKFLPFLTNTIETLNNLVNKIDDDYLLNKKAKINDCNVEKDEAHVKITKLSDLK 395
Qy 278 AIDDKIDLFKNPYDFAEATKLLINDTCKKMLGKLLSTGLVONFPNTIISKLEIEGKFQDML 337
Db 455 AIDDKIDLFKNPYDFAEATKLLINDTCKKMLGKLLSTGLVONFPNTIISKLEIEGKFQDML 455
Qy 338 NISQHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNNGGCCADA 397
Db 456 NISQHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNNGGCCADA 515
Qy 398 TCTEEDSGSSRRKKTCTCTKPDSPYPLFDGIFCSS 431
Db 516 KCTEEDSGSNGKKITCTCTKPDSPYPLFDGIFCSS 549
```

```
Db 336 TDEQKESKFLPFLTNTIETLNNLVNKIDDDYLLNKKAKINDCNVEKDEAHVKITKLSDLK 395
Qy 278 AIDDKIDLFKNPYDFAEATKLLINDTCKKMLGKLLSTGLVONFPNTIISKLEIEGKFQDML 337
Db 455 AIDDKIDLFKNPYDFAEATKLLINDTCKKMLGKLLSTGLVONFPNTIISKLEIEGKFQDML 455
Qy 338 NISQHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNNGGCCADA 397
Db 456 NISQHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNNGGCCADA 515
Qy 398 TCTEEDSGSSRRKKTCTCTKPDSPYPLFDGIFCSS 431
Db 516 KCTEEDSGSNGKKITCTCTKPDSPYPLFDGIFCSS 549

RESULT 27
Q25979 ID Q25979 PRELIMINARY; PRT; 569 AA.
AC Q25979;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13352; BAA02611.1; -
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;

Query Match 84.28; Score 1915; DB 5; Length 569;
Best Local Similarity 93.98; Pred. No. 1.2e-85;
Matches 370; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

Qy 38 TDDDDKAMADIGTSGRGCTMAISVTMDNILSGFENEYDVIYKPLAGYVRSLLKKQIEKNI 97
Db 162 SEDNDEYLDQVVTGE-----AISTVMDNILSGFENEYDVIYKPLAGYVRSLLKKQIEKNI 216
Qy 98 FTNLLNLDILNSRLKRRKYFDVLESQKHHISSNEYIIDSFKLLNSEQKNTLLKS 157
Db 217 ITFNLNLDILNSRLKRRKYFDVLESQKHHISSNEYIIDSFKLLNSEQKNTLLKS 276
Qy 158 YKYIKESVENDIKFAQEGISYEVYKLAQKDDLESIKKVIKEEKEKFPSSPTTTPSPAK 217
Db 277 YKYIKESVENDIKFAQEGISYEVYKLAQKDDLESIKKVIKEEKEKFPSSPTTTPSPAK 335
Qy 218 TDEQKESKFLPFLTNTIETLNNLVNKIDDDYLLNKKAKINDCNVEKDEAHVKITKLSDLK 277
Db 336 TDEQKESKFLPFLTNTIETLNNLVNKIDDDYLLNKKAKINDCNVEKDEAHVKITKLSDLK 395
Qy 278 AIDDKIDLFKNPYDFAEATKLLINDTCKKMLGKLLSTGLVONFPNTIISKLEIEGKFQDML 337
Db 455 AIDDKIDLFKNPYDFAEATKLLINDTCKKMLGKLLSTGLVONFPNTIISKLEIEGKFQDML 455
Qy 338 NISQHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNNGGCCADA 397
Db 456 NISQHCVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNNGGCCADA 515
Qy 398 TCTEEDSGSSRRKKTCTCTKPDSPYPLFDGIFCSS 431
Db 516 KCTEEDSGSNGKKITCTCTKPDSPYPLFDGIFCSS 549
```



```

RESULT 28
Q9NHX1 PRELIMINARY; PRT; 1694 AA.
AC Q9NHX1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Major merozoite surface antigen.
GN GP195.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RC STRAIN=FCCL/HN;
RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.;
RT "Molecular cloning and sequence analysis of major merozoite surface
antigen(gp195)gene of Plasmodium falciparum isolate FCCL/HN.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF218248; AAF27526.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW Merozoite.
SQ SEQUENCE 1694 AA; 192767 MW; B51634A49E0F6728 CRC64;

Query Match 84.2%; Score 1915; DB 5; Length 1694;
Best Local Similarity 93.9%; Pred. No. 3.8e-85;
Matches 370; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

QY 38 TDDDDKAMADIGSIEGRGTMAISVTMDNLSGFENEYDVYIYKPLAGVYRSLLKKQIEKNI 97
Db 1287 SEDNDEYLDQVVTGE-----AISVTMDNLSGFENEYDVYIYKPLAGVYRSLLKKQIEKNI 1341
QY 98 FTFNLNLDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSPKLLNSEQKNTLLKS 157
Db 1342 ITFNLNLDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSPKLLNSEQKNTLLKS 1401
QY 158 YKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 217
Db 1402 YKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 1460
QY 218 TDQKKESKFLPPLTNIETLYNNLVNKIDYDLYNLKAKINDCNVDEKDEAHVKITKLSDLK 277
Db 1461 TDQKKESKFLPPLTNIETLYNNLVNKIDYDLYNLKAKINDCNVDEKDEAHVKITKLSDLK 1520
QY 278 AIDDKIDLFPKNTDFAIKKLINDDTKMDLGLLSTGLVQNPNTIISKLEGGKFDML 337
Db 1521 AIDDKIDLFPKNTDFAIKKLINDDTKMDLGLLSTGLVQNPNTIISKLEGGKFDML 1580
QY 338 NISQHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNTCNENNGGCDADA 397
Db 1581 NISQHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNTCNENNGGCDADA 1640
QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431
Db 1641 KCTEEDSGSGNGKITCECTKPDSPYPLFDGIFCSS 1674

RESULT 29
Q9T2T5 PRELIMINARY; PRT; 1694 AA.
AC Q9T2T5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1.
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RC STRAIN=HN2;
RA Jiang G., Liu R., Daubenberger C.A., Pluschke G.;
RT "Sequence analysis of the MSP 1 gene of Plasmodium falciparum from
Hainan, China.";
RL Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung 17:294-297(1999).
DR EMBL: AF062349; AAC72885.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW Merozoite surface protein 1.
SQ SEQUENCE 1704 AA; 193763 MW; 385526D0DA56FD1D CRC64;

Query Match 84.2%; Score 1915; DB 5; Length 1704;
Best Local Similarity 93.9%; Pred. No. 3.8e-85;
Matches 370; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

QY 38 TDDDDKAMADIGSIEGRGTMAISVTMDNLSGFENEYDVYIYKPLAGVYRSLLKKQIEKNI 97
Db 1287 SEDNDEYLDQVVTGE-----AISVTMDNLSGFENEYDVYIYKPLAGVYRSLLKKQIEKNI 1341
QY 98 FTFNLNLDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSPKLLNSEQKNTLLKS 157
Db 1342 ITFNLNLDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSPKLLNSEQKNTLLKS 1401
QY 158 YKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 217
Db 1402 YKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 1460
QY 218 TDQKKESKFLPPLTNIETLYNNLVNKIDYDLYNLKAKINDCNVDEKDEAHVKITKLSDLK 277
Db 1461 TDQKKESKFLPPLTNIETLYNNLVNKIDYDLYNLKAKINDCNVDEKDEAHVKITKLSDLK 1520
QY 278 AIDDKIDLFPKNTDFAIKKLINDDTKMDLGLLSTGLVQNPNTIISKLEGGKFDML 337
Db 1521 AIDDKIDLFPKNTDFAIKKLINDDTKMDLGLLSTGLVQNPNTIISKLEGGKFDML 1580
QY 338 NISQHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNTCNENNGGCDADA 397
Db 1581 NISQHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNTCNENNGGCDADA 1640
QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431
Db 1641 KCTEEDSGSGNGKITCECTKPDSPYPLFDGIFCSS 1674

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=HN1;
RA Jiang G., Liu R., Daubenberger C.A., Pluschke G.;
RT "Sequence analysis of the MSP 1 gene of Plasmodium falciparum from
Hainan, China.";
RL Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung 17:294-297(1999).
DR EMBL: AF062348; AAC72884.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
SQ SEQUENCE 1694 AA; 192795 MW; 84CFC0E709F5673B CRC64;

Query Match 84.2%; Score 1915; DB 5; Length 1694;
Best Local Similarity 93.9%; Pred. No. 3.8e-85;
Matches 370; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

QY 38 TDDDDKAMADIGSIEGRGTMAISVTMDNLSGFENEYDVYIYKPLAGVYRSLLKKQIEKNI 97
Db 1287 SEDNDEYLDQVVTGE-----AISVTMDNLSGFENEYDVYIYKPLAGVYRSLLKKQIEKNI 1341
QY 98 FTFNLNLDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSPKLLNSEQKNTLLKS 157
Db 1342 ITFNLNLDILNSRLKRRKYFLDVLSDLMQFKHISSEYIIEDSPKLLNSEQKNTLLKS 1401
QY 158 YKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 217
Db 1402 YKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 1460
QY 218 TDQKKESKFLPPLTNIETLYNNLVNKIDYDLYNLKAKINDCNVDEKDEAHVKITKLSDLK 277
Db 1461 TDQKKESKFLPPLTNIETLYNNLVNKIDYDLYNLKAKINDCNVDEKDEAHVKITKLSDLK 1520
QY 278 AIDDKIDLFPKNTDFAIKKLINDDTKMDLGLLSTGLVQNPNTIISKLEGGKFDML 337
Db 1521 AIDDKIDLFPKNTDFAIKKLINDDTKMDLGLLSTGLVQNPNTIISKLEGGKFDML 1580
QY 338 NISQHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNTCNENNGGCDADA 397
Db 1581 NISQHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNTCNENNGGCDADA 1640
QY 398 TCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431
Db 1641 KCTEEDSGSGNGKITCECTKPDSPYPLFDGIFCSS 1674

RESULT 30
Q9T2T4 PRELIMINARY; PRT; 1704 AA.
AC Q9T2T4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1.
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RC STRAIN=HN2;
RA Jiang G., Liu R., Daubenberger C.A., Pluschke G.;
RT "Sequence analysis of the MSP 1 gene of Plasmodium falciparum from
Hainan, China.";
RL Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung 17:294-297(1999).
DR EMBL: AF062349; AAC72885.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
SQ SEQUENCE 1704 AA; 193763 MW; 385526D0DA56FD1D CRC64;

Query Match 84.2%; Score 1915; DB 5; Length 1704;
Best Local Similarity 93.9%; Pred. No. 3.8e-85;
Matches 370; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

QY 38 TDDDDKAMADIGSIEGRGTMAISVTMDNLSGFENEYDVYIYKPLAGVYRSLLKKQIEKNI 97

```

1297	SENDEYLDQVVTGE-----AISTVMDNILSGFENEVDYILKPLAGYYLSLKKQIEKNI	1351
98	FTFNLNLDILNSRLKRRKYFDVLESDLQFKHISISNEYIIEDSPFKLLNSEQKNTLLKS	157
1352	ITFNLNLDILNSRLKRRKYFDVLESDLQFKHISISNEYIIEDSPFKLLNSEQKNTLLKS	1411
158	YKYIKESVENDIKFPAQEGISYIEKYVLAKYKDDLESIKVKYKEEKEKTPSPPTTSPSPAK	217
1412	YKYIKESVENDIKFPAQEGISYIEKYVLAKYKDDLESIKVKYKEEKE-EPSPPTTSPSPAK	1470
218	TDEQKESKFLPFLNIETLYNNLVNKIDDDYILNLKAKINDCNVKEDEAHVTKKLSDLK	277
1471	TDEQKESKFLPFLNIETLYNNLVNKIDDDYILNLKAKINDCNVKEDEAHVTKKLSDLK	1530
278	AIDDKIDLKFNYPDEFAIKKLLINDDTTKKMDLKGKLLSTGLVQNFPTTIISKLIEGKQDML	337
1531	AIDDKIDLKFNNTDSEAIKKLLINDDTTKKMDLKGKLLSTGLVQNFPTTIISKLIEGKQDML	1590
338	NTSQCVCVKKQCPENSGGFRHLDEREECKCLLNTKQEGDKCVENPNPTCNENNGGCCDADA	397
1591	NTSQCVCVKKQCPENSGGFRHLDEREECKCLLNTKQEGDKCVENPNPTCNENNGGCCDADA	1650
398	TCTEEDSGSRKKKITCECTKPDSPYLPDGFICSS	431
1651	KCTEEDSGNGKKKITCECTKPDSPYLPDGFICSS	1684

```

RESULT 31
Q25725
ID Q25725 PRELIMINARY; PRT; 372 AA..
AC Q25725;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5833;
RN [1]
RC
RP SEQUENCE FROM N.A.
RC STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
RL (MSP-1) in field-derived Plasmodium falciparum."
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RL EMBL; U20730; AA62221.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1
FT NON_TER 372
SQ SEQUENCE 372 AA; 42711 MW; E3DBB33F282CD5FC CRC64;

```

[illegible]

```

RESULT 32
Q25726 ID Q25726 PRELIMINARY; PRT; 372 AA.
AC Q25726;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Pova M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1 (MSP-1) in field-derived Plasmodium falciparum."
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20731; AAA6222.1;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1
DR NON_TER 1
ET NON_TER 372 372
SQ SEQUENCE 372 AA; 42734 MW; 19844F60D7011E52 CRC64;

```

Query Match	83.6%; Score 1901.5; DB 5; Length 372;
Best Local Similarity	97.9%; Pred. No. 3.5e-85;
Matches 365; Conservative 3; Mismatches 4; Indels 1; Gaps 1;	
Qy	58 AISYTMNDILSGFNEVDVIYKPLAGYRSLKKQIEKNFTFNLNLDILNSRLKKRKY 117
Db	1 AISYTMDBHILSGFNEVDVIYKPLAGYRSLKKQIEKNFTFNLNLDILNSRLKKRKY 60
Qy	118 FLDVLESDLMOFKHISSNEYIIEDSFLLNSEQNTLLKSYKYIKESVENDIKFQAEGIS 177
Db	61 FLDVLESDLMOFKHISSNEYIIEDSFLLNSEQNTLLKSYKYIKESVENDIKFQAEGIS 120
Qy	178 YYEKVLAKYKDDLESIKVKIEKKEKPPSPPTPPSPAKTDEOKKESKFLPFLTNITEL 237
Db	121 YYEKVLAKYKDDLESIKVKIEKEE-PPSSPTTTPPSPAKTDEOKKESKFLPFLTNITEL 179
Qy	238 YNNLVNKIDDIYLNILKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNPYDFEAIKK 297
Db	180 YNNLVNKIDDIYLNILKAKINDCNVKEKNEAHVKITKLSDLKAIDDKIDLFKNHNDFEAIKK 239
Qy	298 LINDDTTKMDLGLKLLSLGLVQNFPNTIISKLIEGKFQDMLNISOHCVKKQCPENSGCFR 357
Db	240 LINDDTTKMDLGLKLLSLGLVQNFPNTIISKLIEGKFQDMLNISOHCVKKQCPENSGCFR 299
Qy	358 HLDERECKCLLNYKQEGDKCVENPNPTCNENNGCCDADATCTEEDSGSRKKKITCCECTK 417
Db	300 HLDERECKCLLNYKQEGDKCVENPNPTCNENNGCCDADAKCTEEDSGSNKKKITCCECTK 359
Qy	418 PDSYPLFDGIFCS 430
Db	360 PDSYPLFDGIFCS 372

KW EGF-like domain; Merozoite.

FT NON_TER 1
SQ SEQUENCE 539 AA; 61076 MW; 72186F1412C57D1C CRC64;

Query Match 51.8%; Score 1178; DB 5; Length 539;

Best Local Similarity 56.9%; Pred. No. 6.3e-50; Indels 24; Gaps 4;
Matches 227; Conservative 58; Mismatches 90;

Qy 34 PDLGTDGDDKAMADIGSTEGRTMAISVTMDNLSGFENEYDVLYLPLAGVYRSLKKQI 93

Db 144 PIFGESEED--YDDLQGVVTEAVTPSV-IDNLSKIENEYEVLYLPLAGVYRSLKKQL 200

Qy 94 EKNIFTNLDNLINSLRKKRYFLDVLSDLMQFKHSSNEYIEDSFLLNSEQKNT 153

Db 201 ENNVMTNVNVKIDILNSRFNFKNVLESDLPYKDLTSSNVVDPYKFLNKRKDK 260

Qy 154 LLKSYKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVIKEEKFPSPPTTPP 213

Db 261 FLSSYNIKDSIDTDINFANDVLGYKILSEKYSDDLDSIKKYI----- 304

Qy 214 SPKATDEQKESKFLPFLTNIETLYNNLVNKIDYDLINLKAINDCNVEKDEAHVKITKL 273

Db 305 ---NDKQGENEKVLPFLNNIETLYKTVDKIDLVFVHLEAKVLNITYEKSNEVEKIKEL 360

Qy 274 SLDKAIDDKIDLFKNPYDFAEIKKLLINDDTTKDMLGKLLSTGLV-QNFPNTIISKLEGG 332

Db 361 NYLKTIOQKLAADFKNFNNFVGIADLSTDYNNHLLTKFLSTGMVFENLAKTVLSNLDGN 420

Qy 333 FQDMLNISOHCVKKQCPENSGCFRHLDEREECKLLNYKQEGDKCVENPNTCNEENGG 392

Db 421 LQGLMNLISOHCVKKQCPENSGCFRHLDEREECKLLNYKQEGDKCVENPNTCNEENGG 480

Qy 393 CDADATCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431

Db 481 CDADAKCTEEDSGSGKKITCECTKPDSPYPLFDGIFCSS 519

RESULT 36

Q25981 ID Q25981 PRELIMINARY; PRT; 539 AA.

AC Q25981;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 01, Last annotation update)

DE Major merozoite surface protein (Fragment).

GN MSP1.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93295445; PubMed=8515786;

RA Jongwutives S., Tanabe K., Kanbara H.;

RT "Sequence conservation in the C-terminal part of the precursor to the

major merozoite surface proteins (MSP1) of Plasmodium falciparum from

field isolates.;

RL Mol. Biochem. Parasitol. 59:95-100(1993).

DR EMBL; D13363; BAA02624.1; -

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR001245; Tyr_pkinase.

DR Pfam; PF00008; EGF; 1

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

KW EGF-like domain; Merozoite.

FT NON_TER 1

SQ SEQUENCE 539 AA; 61046 MW; 3EED87473EE87B65 CRC64;

Query Match 51.8%; Score 1178; DB 5; Length 539;

Best Local Similarity 56.9%; Pred. No. 6.3e-50;

Matches 227; Conservative 58; Mismatches 90; Indels 24; Gaps 4;

Qy 34 PDLGTDGDDKAMADIGSTEGRTMAISVTMDNLSGFENEYDVLYLPLAGVYRSLKKQI 93

Db 144 PIFGESEED--YDDLQGVVTEAVTPSV-IDNLSKIENEYEVLYLPLAGVYRSLKKQL 200

Qy 94 EKNIFTNLDNLINSLRKKRYFLDVLSDLMQFKHSSNEYIEDSFLLNSEQKNT 153

Db 201 ENNVMTNVNVKIDILNSRFNFKNVLESDLPYKDLTSSNVVDPYKFLNKRKDK 260

Qy 154 LLKSYKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVIKEEKFPSPPTTPP 213

Db 261 FLSSYNIKDSIDTDINFANDVLGYKILSEKYSDDLDSIKKYI----- 304

Qy 214 SPKATDEQKESKFLPFLTNIETLYNNLVNKIDYDLINLKAINDCNVEKDEAHVKITKL 273

Db 305 ---NDKQGENEKVLPFLNNIETLYKTVDKIDLVFVHLEAKVLNITYEKSNEVEKIKEL 360

Qy 274 SLDKAIDDKIDLFKNPYDFAEIKKLLINDDTTKDMLGKLLSTGLV-QNFPNTIISKLEGG 332

Db 361 NYLKTIOQKLAADFKNFNNFVGIADLSTDYNNHLLTKFLSTGMVFENLAKTVLSNLDGN 420

Qy 333 FQDMLNISOHCVKKQCPENSGCFRHLDEREECKLLNYKQEGDKCVENPNTCNEENGG 392

Db 421 LQGLMNLISOHCVKKQCPENSGCFRHLDEREECKLLNYKQEGDKCVENPNTCNEENGG 480

Qy 393 CDADATCTEEDSGSRKKITCECTKPDSPYPLFDGIFCSS 431

Db 481 CDADAKCTEEDSGSGKKITCECTKPDSPYPLFDGIFCSS 519

RESULT 37

Q03999 ID Q03999 PRELIMINARY; PRT; 400 AA.

AC Q03999;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE Merozoite surface protein 1 precursor (Fragment).

GN MSP1.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

RP SEQUENCE FROM N.A.

RA Blackman M.J., Ling I.T., Nicholls S.C., Holder A.A.;

RT "Proteolytic processing of the Plasmodium falciparum merozoite surface

protein-1 produces a membrane-bound fragment containing two epidermal

growth factor-like domains.;

RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.

DR EMBL; M64681; AAA29709.1; -

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR001245; Tyr_pkinase.

DR Pfam; PF00008; EGF; 1.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

KW Signal.

FT NON_TER 1

FT SIGNAL <1 24 POTENTIAL.

FT CHAIN 25 286 MEROZOITE SURFACE PROTEIN 1.

FT CHAIN 287 400 MEROZOITE SURFACE PROTEIN 1.

SQ SEQUENCE 400 AA; 45824 MW; 537F075058626AC2 CRC64;

Query Match 51.6%; Score 1175; DB 5; Length 400;

Best Local Similarity 56.6%; Pred. No. 6.4e-50;

Matches 226; Conservative 59; Mismatches 90; Indels +24; Gaps 4;

Qy 34 PDLGTDGDDKAMADIGSTEGRTMAISVTMDNLSGFENEYDVLYLPLAGVYRSLKKQI 93

Db 5 PIFGESEED--YDDLQGVVTEAVTPSV-IDNLSKIENEYEVLYLPLAGVYRSLKKQL 61

Qy 94 EKNIFTNLDNLINSLRKKRYFLDVLSDLMQFKHSSNEYIEDSFLLNSEQKNT 153

Db 62 ENNVMTNVNVKIDILNSRFNFKNVLESDLPYKDLTSSNVVDPYKFLNKRKDK 121

Qy 154 LLKSYKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVIKEEKFPSPPTTPP 213

Db 122 FLSSYNIKDSIDTDINFANDVLGYKILSEKYSDDLDSIKKYI----- 165

```
QY 214 SPAKTDQKKESKFLPFLNIETLYNNLVNKIDDDYILNKLKAKINDCNVEKDEAHVKITKL 273
Db 166 ----NDKQGENEKYLPFLNIETLYKTVDNKIDLFVHLEAKVNLNTYKSNVEVKIKEL 221
QY 274 SOLKATDDKIDLFKNPYDFAEAKKLLNDTKKMDLGLKLLSTGLV-QNFPNTIISKIEGK 332
Db 222 NYLKTIDQKLADFKNKNFVGIADLSTDYNNHNNLLTKFLSTGMVFENLAKTVLSNLLDGN 281
QY 333 FQMLNISQHCQVKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGG 392
Db 282 LOGMLNISQHCQVKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGG 341
QY 393 CDADATCTEEDSGSRKKITCTCKPDSYPLFDGIFCSS 431
Db 342 CDADAKCTEEDSGSGNKKITCTCKPDSYPLFDGIFCSS 380

RESULT 38
Q25984 ID Q25984 PRELIMINARY; PRT; 539 AA.
AC Q25984;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Major merozoite surface protein (fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPI) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13356; BAA02617.1;
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61115 MW; 3788015F3127CB9E CRC64;

Query Match 51.6%; Score 1175; DB 5; Length 539;
Best Local Similarity 56.6%; Pred. No. 8.8e-50;
Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;

QY 34 PDLTDDDDKAMADIGSIEGRGTMAISVTMDNITLSGFENEYDVIYKPLAGVYRSLLKQI 93
Db 144 PIFGESEED--YDGLQVVTGEAVTPSV-IDNLSKIENEYEVLYLKPLAGVYRSLLKQI 200
QY 94 EKNIFTNLDNLDILNSRLKKRYKFLDVLESMDLQMFKHSSNEYIIEDSFKLNSQKNT 153
Db 201 ENNVMTFNVNVDILNSRENKRENFKNVLESOLIPYKOLTSNNYVVKDPYKFLNKEKRD 260
QY 154 LKSYKIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTP 213
Db 261 FLSSYNYIKDSIDTDFNANDVLGVYKILSEKYSOLDSIKKI----- 304
QY 214 SPAKTDQKKESKFLPFLNIETLYNNLVNKIDDDYILNKLKAKINDCNVEKDEAHVKITKL 273
Db 305 ----NDKQGENEKYLPFLNIETLYKTVDNKIDLFVHLEAKVNLNTYKSNVEVKIKEL 360
QY 274 SOLKATDDKIDLFKNPYDFAEAKKLLNDTKKMDLGLKLLSTGLV-QNFPNTIISKIEGK 332
Db 361 NYLKTIDQKLADFKNKNFVGIADLSTDYNNHNNLLTKFLSTGMVFENLAKTVLSNLLDGN 420
QY 333 FQMLNISQHCQVKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGG 392
Db 421 LOGMLNISQHCQVKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGG 480

RESULT 40
Q25984 ID Q25984 PRELIMINARY; PRT; 539 AA.
AC Q25984;
DT 01-MAY-2000 (TremBLrel. 13, Created)
```

```
QY 393 CDADATCTEEDSGSRKKITCTCKPDSYPLFDGIFCSS 431
Db 481 CDADAKCTEEDSGSGNKKITCTCKPDSYPLFDGIFCSS 519

RESULT 39
Q25966 ID Q25966 PRELIMINARY; PRT; 539 AA.
AC Q25966;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Major merozoite surface protein (fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPI) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13357; BAA02618.1;
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61145 MW; 7B7DE90C1D0ACDE7 CRC64;

Query Match 51.6%; Score 1175; DB 5; Length 539;
Best Local Similarity 56.6%; Pred. No. 8.8e-50;
Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;

QY 34 PDLTDDDDKAMADIGSIEGRGTMAISVTMDNITLSGFENEYDVIYKPLAGVYRSLLKQI 93
Db 144 PIFGESEED--YDGLQVVTGEAVTPSV-IDNLSKIENEYEVLYLKPLAGVYRSLLKQI 200
QY 94 EKNIFTNLDNLDILNSRLKKRYKFLDVLESMDLQMFKHSSNEYIIEDSFKLNSQKNT 153
Db 201 ENNVMTFNVNVDILNSRENKRENFKNVLESOLIPYKOLTSNNYVVKDPYKFLNKEKRD 260
QY 154 LKSYKIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEEKEKFPSSPTTP 213
Db 261 FLSSYNYIKDSIDTDFNANDVLGVYKILSEKYSOLDSIKKI----- 304
QY 214 SPAKTDQKKESKFLPFLNIETLYNNLVNKIDDDYILNKLKAKINDCNVEKDEAHVKITKL 273
Db 305 ----NDKQGENEKYLPFLNIETLYKTVDNKIDLFVHLEAKVNLNTYKSNVEVKIKEL 360
QY 274 SOLKATDDKIDLFKNPYDFAEAKKLLNDTKKMDLGLKLLSTGLV-QNFPNTIISKIEGK 332
Db 361 NYLKTIDQKLADFKNKNFVGIADLSTDYNNHNNLLTKFLSTGMVFENLAKTVLSNLLDGN 420
QY 333 FQMLNISQHCQVKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGG 392
Db 421 LOGMLNISQHCQVKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGG 480

RESULT 40
Q25966 ID Q25966 PRELIMINARY; PRT; 539 AA.
AC Q25966;
DT 01-MAY-2000 (TremBLrel. 13, Created)
```

01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
Major merozoite surface protein (Fragment).
MSPl.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
major merozoite surface proteins (MSPl) of Plasmodium falciparum from
field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13358; BAA02619.1; -;
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001245; Tyr-kinase.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61075 MW; C70C2E100EC4A101 CRC64;

Query Match 51.6%; Score 1175; DB 5; Length 539;
Best Local Similarity 56.6%; Pred. No. 8.8e-50;
Matches 226; Conservative 59; Mismatches 90; Indels 24; Gaps 4;

QY 34 PDLGTDGDDKAMADIGSTEGRTMAISVTMDNIIILSGFENEYDVYIPLAGVYRSLKKOI 93
DB 144 PIFGESED--YDILGQVVTGEAVTPSV-IDNIIILSKENEYEVYIPLAGVYRSLKKOL 200

QY 94 EKNIFTNLNLDILNSRLKRYFLDVLDESILMQFKHISSNEYIIEDSEKLLNSEQKNT 153
DB 201 ENNVMTENVVVKDILNSRFNKNFNKLVESDLIPYKDLTSSNVVVKDPYKFLNKRDK 260

QY 154 LLKSYKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVIKEKEKPPSPPTTP 213
DB 261 FLSSYNYIKDSIDTINFANDVLGYKILSEKYSLDLSIKKYI----- 304

QY 214 SPAKTDEQKESKFLPFTNLTETLYNLVKNIDDDYILNKLAKINDCNVKEDEAHVKITKL 273
DB 305 ----NDKQGENEKYLPFNNTIETLYKTVNDKIDLFVHLEAKVLYNYEKSNNVEVKIKEL 360

QY 274 SLDKAIIDDKIDLFKNPYDFAEIKKILINDTKKMLGKLLSTGLV-QNPPNTIISKLIIEGK 332
DB 361 NYLKTIQDKLADFKNNFNFGIADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLLDGN 420

QY 333 FODMLNISQHCYKQCPENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGG 392
DB 421 LQGLMLNISQHCYKQCPQNSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGG 480

QY 393 CDADATCTEEDSGSRKKITCECTKPDSPFLFDGIFCSS 431
DB 481 CDADAKTEEDSGSGNGKKTCECTKPDSPFLFDGIFCSS 519

Search completed: May 12, 2003, 10:19:20
Job time : 95 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 19, 2003, 14:37:01 ; Search time 2567 Seconds
(without alignments)
4886.369 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 2275

Sequence: 1 MHHHHHSSGLVPRGSMK.....TCECTKPSYPLFDGIFCSS 431

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.scool/US10057531/runat_12052003.091031_21666/app_query.fasta_1.583
-DB=GenEmbl -QFT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10057531.ecgn.1.1.1687 -runat_12052003.091031_21666 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-FGAPOPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	1970	86.6	1249	12	AF395808	AF395808 Synthetic
2	1969.5	86.6	5312	3	PFMEZSA1B	Z35327 P.falciiparu
3	1965.5	86.4	1959	3	PFMEZSA1C	Z35328 P.falciiparu
4	1946.5	85.6	1729	3	PFAMSP814	D13345 P. falciipar
5	1946.5	85.6	1729	3	PFAMSPM20	D13343 P. falciipar
6	1946.5	85.6	5392	3	PF190G1	X05824 Plasmodium
7	1946.5	85.6	5754	3	PFAMMSAA	M19143 P.falciiparu
8	1942	85.4	1119	3	PFU20728	U20728 Plasmodium
9	1937	85.1	1726	3	PFAMSP841	D13351 P. falciipar
10	1935.5	85.1	5541	3	PFPG195	X03831 P.falciiparu
11	1935	85.1	1119	3	PFU20727	U20727 Plasmodium
12	1935	85.1	1956	3	PFMEZSA1A	Z35326 P.falciiparu
13	1933	85.0	1119	3	PFU20732	U20732 Plasmodium
14	1932.5	84.9	5276	3	PFAGP195	M37213 P.falciiparu
15	1931	84.9	1726	3	PFAMSP8221	D13347 P. falciipar
16	1931	84.9	1726	3	PFAMSP843	D13353 P. falciipar
17	1931	84.9	1726	3	PFAMSP946	D13354 P. falciipar
18	1929	84.8	1119	3	AF040568	AF040568 Plasmodiu
19	1928	84.7	1119	3	PFU20733	U20733 Plasmodium
20	1924	84.6	1119	3	PFU20729	U20729 Plasmodium
21	1922.5	84.5	1116	3	PFU20653	U20653 Plasmodium
22	1921	84.4	1726	3	PFAMSP947	D13355 P. falciipar
23	1919	84.4	1119	3	AF040567	AF040567 Plasmodiu
24	1919	84.4	1119	3	PFU20726	U20726 Plasmodium
25	1917.5	84.3	1116	3	PFU20654	U20654 Plasmodium
26	1917.5	84.3	1116	3	PFU20655	U20655 Plasmodium
27	1917.5	84.3	1116	3	PFU20656	U20656 Plasmodium
28	1917	84.3	1956	3	PFMEZSA1D	D13346 P. falciipar
29	1915	84.2	1726	3	PFAMSP815	D13346 P. falciipar
30	1915	84.2	1726	3	PFAMSP8351	D13348 P. falciipar
31	1915	84.2	1726	3	PFAMSP8352	D13349 P. falciipar
32	1915	84.2	1726	3	PFAMSP838	D13350 P. falciipar
33	1915	84.2	1726	3	PFAMSP842	D13352 P. falciipar
34	1915	84.2	5085	3	AF218248	AF218248 Plasmodiu
35	1915	84.2	5243	3	AF062348	AF062348 Plasmodiu
36	1915	84.2	5271	3	AF062349	AF062349 Plasmodiu
37	1906.5	83.8	1116	3	PFU20730	U20730 Plasmodium
38	1901.5	83.6	1116	3	PFU20731	U20731 Plasmodium
39	1900.5	83.5	1116	3	AF040569	AF040569 Plasmodiu
40	1900	83.5	1726	3	PFAMSP807	D13344 P. falciipar
41	1178	51.8	1636	3	PFAMSP828	D13360 P. falciipar
42	1178	51.8	1636	3	PFAMSP844	D13363 P. falciipar
43	1175	51.6	1203	3	PFAMSP1	M64681 Plasmodium
44	1175	51.6	1636	3	PFAMSP806	D13357 P. falciipar
45	1175	51.6	1636	3	PFAMSP808	D13358 P. falciipar

ALIGNMENTS

RESULT 1

AF395808 1249 bp mRNA linear SYN 20-AUG-2001
 LOCUS Synthetic construct melittin signal peptide/MSPl-42 fusion protein
 DEFINITION mRNA, complete cds.
 ACCESSION AF395808
 VERSION AF395808.1 GI:15213639
 KEYWORDS synthetic construct.
 SOURCE synthetic construct
 ORGANISM artificial sequences.
 REFERENCE 1 (bases 1 to 1249)
 AUTHORS Pang,L.Y., Hashimoto,C.N., Tam,L.Q., Meng,Z.Q., Hui,S.N. and
 Ho,K.K.
 TITLE In vivo expression and immunological studies of the 42 kDa
 carboxyl-terminal processing fragment of Plasmodium falciparum
 merozoite surface protein 1 (MSPl-42) in the Baculovirus-silkworm
 system
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1249)
 AUTHORS Pang,L.Y., Hashimoto,C.N., Tam,L.Q., Meng,Z.Q., Hui,S.N. and
 Ho,K.K.
 TITLE Direct Submission
 JOURNAL Submitted (28-JUN-2001) Department of Biochemistry, The Chinese
 University of Hong Kong, Shatin, Hong Kong
 FEATURES
 source
 1..1249
 /organism="synthetic construct"
 /isolate="3D7"
 /db_xref="taxon:32630"
 26..1228
 /note="honeybee melittin signal sequence fused with 42 kDa
 carboxyl-terminal processing fragment of Plasmodium
 falciparum isolate 3D7 merozoite surface protein 1"
 /codon_start=1
 /product="melittin signal peptide/MSPl-42 fusion protein"
 /protein_id="AAK92098.1"
 /db_xref="GI:15213640"
 /translation="MFLVNVALVFMVYVYISYIADPSPMAISVTMDNLSGFENEYD
 VVILPLAGYVRLSKQIEKNIETFLNLDNLSRLKRYFLDVLSDLMQFKHIS
 SNEYIDSKLNSKQNTLLSKYIKESVENDIKFAQEGISYEKVLAKYKDDLE
 SIKKVIKEKEKPPSPPTPPSPAKTDEQKESKLPFLTNIETLYNNLVNKIDYDL
 INLAKINDCNVEKDEAHVKITLSDLKAIDDKIDLPFNYPDEAKKLINDITKDKM
 IGLLLTGLVGNFPNTIISKLIKGKFDMLNISOHQVCKQCQKQCPENSGCFRHLDEREC
 KCLNLYKQEGDKGVENPNPTNNNGGCDATCTEEDSGSRKKIKTCETKPKDSYPL
 FDGIFCSS"
 BASE COUNT 518 a 180 c 185 g 366 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.62e-129 Length: 1249
 Score: 1970.00 Matches: 375
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 86.59% Indels: 0
 DB: 12 Gaps: 0
 US-10-057-531a-2 (1-431) x AF395808 (1-1249)
 Qy 57 MetAlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspVal 76
 Db 101 ATGGCAATATCTGCACAAATGGTAATATCTCTCAGGATTTGAAATATGATATGTTT 160
 Qy 77 IleTyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsn 96
 Db 161 ATATATTAAACCTTTAGCTGGAGTATATAGACAGCTTAAAAAACAATTTGAAAAAAC 220
 Qy 97 IlePheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLys 116
 Db 221 ATTTTACATTAATTTAAATTTGAACGATATCTTAAATTCACGCTCTTAAGAAACGAAAA 280
 Qy 117 TyrPheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGlu 136

Db 281 TATTCTTAGATGTATTAGAAATCTGATTAAATGAATTTAAACATATATCTCAATGAA 340
 Qy 137 TyrIleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLys 156
 Db 341 TACATTATTGAAGATTCAATTTAAATTTATGATTTCAGAACAAAAAACACACTTTTAAAA 400
 Qy 157 SerTyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIle 176
 Db 401 AGTTACAAATATATAAAGAAATCAGTACAAATGATATTAAATTTGCACAGGAAGGTATA 460
 Qy 177 SerTyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysVal 196
 Db 461 AGTTATTATGAAAGGTTTATAGCGAAATATAGAGTATTTAGAAATCAATTTAAAAAGTT 520
 Qy 197 IleLysGluGluLysGluLysPheProSerProProThrThrProProSerProAla 216
 Db 521 ATCAAGAGAGAGAGAGAGAGTTCCCATCATCACCAACCAACACCTCCCTCCACGACA 580
 Qy 217 LysThrAspGluGlnLysLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThr 236
 Db 581 AAAACAGACGACAAACAAAGAGAAAGTAAGTTCTCTTCCATTTTAAACAAACATTGAGACC 640
 Qy 237 LeuTyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIle 256
 Db 641 TTATACAATAACTTAGTTAATAAATTTGACGATTACTTAATTAATTAACGCAAGGAAAGATT 700
 Qy 257 AsnAspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeu 276
 Db 701 AACGATTGTATGTGTGAAAGAGATGAACACATGTTAAATTAACATACTTAGTGATTATA 760
 Qy 277 LysAlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLys 296
 Db 761 AAGCAATTTGATGACAAATAGATCTTTTAAAAACCCCTTACGACTTCCGAAGCAATTAATA 820
 Qy 297 LysLeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeu 316
 Db 821 AATTTGATAATGATGATACCAAAAAAGATATGCTTGGCAAAATTTACTTAGTACAGGATTA 880
 Qy 317 ValGlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMet 336
 Db 881 GTTCAAAATTTTCTCTAATACATAATATCAAAATTAATTTGAAGGAAAAATTTCCAAGATATG 940
 Qy 337 LeuAsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPhe 356
 Db 941 TTAACATTTTCCACAAACCAATGCGTAAAAAACAATGTCAGAAAAATTTCTGGATGTTTC 1000
 Qy 357 ArgHisLeuAspGluLysGluGluCysLysCysLysLeuLeuAsnTyrLysGlnGluLysAsp 376
 Db 1001 AGACATTTTAGATGAAGAGAGAAATGTAATGTTTATTAATTTACAAACAAGAGGTGAT 1060
 Qy 377 LysCysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAsp 396
 Db 1061 AATGTGTTGAAATCCAAATCCTACTTGTAAACGAAAAATATGTTGGATGTCAGAT 1120
 Qy 397 AlaThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThr 416
 Db 1121 GCCACATGTACCGAAGAGATTCAGGTAGCAGCAGAGAAAAATCACAATGTAATGTACT 1180
 Qy 417 LysProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
 Db 1181 AAACCTGATCTTATCCACTTTTCGATGGTATTTTCTGCAGTTCC 1225

RESULT 2

PFMEZSA1B

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

PFMEZSA1B 5312 bp DNA linear INV 21-FEB-2002
 P.falciparum gpl190 (MSAL, MSPL, PMWSA) for precursor of major
 merozoite surface antigens.

235327
 gpl190 gene; merozoite surface antigen; MSAL gene; MSPL gene; PMWSA
 gene.

malaria parasite P. falciparum.
 Plasmodium falciparum


```

Db 4927 ACATTTTCACACCAATCGTAAACAAATGTCAGAAATTTCTGGATGTTTCAGA 4986
QY 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
Db 4987 CATTTAGATGAAAGACGAATGTAATCTTATTAAATACAAACAAAGGTGATAA 5046
QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyCysAspAlaAspAla 397
Db 5047 TGTGTTGAAATCCAAATCCTACTTGTAAACGAAATTAATGGTGGATGTCAGATGCC 5106
QY 398 ThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys 417
Db 5107 ACATGTACCAGAGAGATTGAGTACGACAGCAAGAAATACATGTCGAATGTACTAAA 5166
QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
Db 5167 CCTGATCTTATCCACTTTTCGATGGTATTTTCTGCAGTTC 5208

RESULT 3
PFMEZSAIC 1959 bp DNA linear INV 30-OCT-1995
P.falciiparum (NF7) gene for merozoite surface antigen 1.
ACCESSION 235328 M19144
VERSION 235328.1 GI:535251
KEYWORDS merozoite surface antigen 1 gene.
SOURCE Plasmodium falciparum.
ORGANISM Plasmodium falciparum.
REFERENCE Peterson,M.G., Coppel,R.L., McIntyre,P., Langford,C.J., Woodrow,G., Brown,G.V., Anders,R.F. and Kemp,D.J. Eukaryota; Alveolata: Apicomplexa; Haemosporida; Plasmodium. Variation in the precursor to the major merozoite surface antigens of Plasmodium falciparum Mol. Biochem. Parasitol. 27 (2-3), 291-301 (1988)
JOURNAL 88142999
MEDLINE 2449612
PUBMED 2 (bases 1 to 1959)
AUTHORS Tolle,R.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-1994) Ralf Tolle, Abt. Prof. H. Bujard, Zentrum fur Molekulare Biologie, (ZMBH), Univ. Heidelberg, Im Neuenheimer Feld 282, Heidelberg, 69120, Germany
AUTHORS 3 (bases 1 to 1959)
REFERENCE Tolle,R., Bujard,H. and Cooper,J.A. Plasmodium falciparum: variations within the C-terminal region of merozoite surface antigen-1 Exp. Parasitol. 81 (1), 47-54 (1995)
JOURNAL 95354793
MEDLINE 7628566
PUBMED 1. .1959
FEATURES Location/Qualifiers
1. .1959
/organism="Plasmodium falciparum"
/isolate="FCH5/NF7"
/db_xref="taxon:5833"
/chromosome="9"
/dev_stage="blood stage"
1. .1959
/gene="gp190, MSAL, MSP1, PMMSA"
<1. .1959
/gene="gp190, MSAL, MSP1, PMMSA"
/codon_start=1
/product="merozoite surface antigen 1 gene, 3+-end"
/protein_id="CAA84557.1"
/db_xref="GI:535252"
/db_xref="SPRMBL:Q25923"
/translation="FNKKKELGQDKMOIKITLLKLEKSLNKNPHNVLFNSVF
FNKKKEIAETENTLENTKILKHYGLVYKNGESSPLKTISEVQIEDNYANLE
KFRVLKIDGLNLDHKKKUSFLSSGLHLITELKEVIKNNKNNPSPSNNKKV
NEALKVSENFLEAKYTVVTPQPPDPSLVRSVSGSGSKTEETQIPTSGSLTE
LQQVQSQDEEDSVLPFIPEGSENDDEYLDQVVTGSAISVTMDNLSGFENEYD
VIVLKLPGAVYRSLLKQIEKNIFTNLDNLNLDLNRKRYFLDVLDESILDMQFKHS
SNEVIIESDFKLNSEQKNTLLSKYIKESVENDIKFAOEGISYYEKLAKYKDDLE

```

```

SIKKVKEKEKFPSPPTTPSPAKTDEQKESKFLPFLTNITLYNNLVNKIDYLL
INLAKINDCNVEKPDIAHVKITKLSDLKIDDKIDLFKNPKYDFEAIKKRLINDTKDM
LGKLLSTGLVONFNTIISKLTIEGFOMLNISOHCVKKOCPEKSGRHLDERREC
KCLLNYKQEGDKCVENPNTCNENNGCGDADATCTEKDSGSRKKITCTEYKPDSPYL
PDGFCSSSNFLGISFLILMLILYSFI"
BASE COUNT 855 a 266 g 573 t
ORIGIN

```

Alignment Scores:

```

Pred. No.: 5, 37e-129 Length: 1959
Score: 1965.50 Matches: 376
Percent Similarity: 97.46% Conservative: 8
Best Local Similarity: 95.43% Mismatches: 5
Query Match: 86.40% Indels: 5
DB: 3 Gaps: 1

US-10-057-531a-2 (1-431) x PFMEZSAIC (1-1959)

QY 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57
Db 730 TCCGAAGATAATGACGAATATTTAGATCAAGTAGTAACGTGGAGAA----- 774
QY 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
Db 775 GCAATATCTGTCCAAATGGATAATATCTCTCAGGATTTGAAATGAATATGATGTTATA 834
QY 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
Db 835 TATTTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAACAATTTGAAAAACAATTT 894
QY 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 895 TTTACATTTAATTTAAATTTGAACGATATCTTAAATTCACGCTTTAAGAAACGAAATAT 954
QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
Db 955 TTTCTAGATGATTATAGAAATCTGATTTAATGCAATTTAAACATATATATCTCAATGAATAC 1014
QY 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
Db 1015 ATTATTGAAGATTCATTTAAATTTATTGAATTCAGAACAAAAACACACTTTTAAAGT 1074
QY 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
Db 1075 TACAAATATATAAAAGATTCAGTAAATATGATATAATTTTGCACAGAGGATTAAGT 1134
QY 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle 197
Db 1135 TATTATGAAAGGTTTTCAGCGAATATATAGGATGATTTAGAAATCAATTTAAAAAGTTATC 1194
QY 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
Db 1195 AAAGAAGAAAAAGGAGAGAGTTCCCATCATCACCAACCAACACCTCCGTCACCACAAAA 1254
QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
Db 1255 ACAGACGAAACAAAAAGAGAAAGTAAGTTCCCTTCCATTTTAAACAAACATTTGAGACCTTA 1314
QY 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
Db 1315 TACAATACTTAGTTAATAAATGACCATTTACTTAATTAATTAAGCGAAAGATTAAAC 1374
QY 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
Db 1375 GATTGTAATGTTGAAAAAGATGAAGCACATGTTTAAATACTAACTTAGTGATTAAAA 1434
QY 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
Db 1435 GCAATTGATGACAAATAGATCTTTTAAAAACCCCTTACGACTTCGAAAGCAATTTAAAA 1494
QY 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
Db 1495 TTGATAAATGATGATACGAAAAAAGATATGCTTGGCAAAATTAATCTAGTACAGGATTAGT 1554

```

QY 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
 DB 1555 CAAATTTTCCCTAATACATATATCAAAATTAATTGAAGGAAATTCACAGATATGTA 1614
 QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
 DB 1615 AACATTTTCAACACACCAATCGCTAAACAAATGTCGAGAAATTCCTGGATGTTTCA 1674
 QY 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
 DB 1675 CATTTAGATGAAGAGAGAAATGTAATGTTTATTAATACAAAGAAAGGTGATAA 1734
 QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397
 DB 1735 TGTGTTGAAATCCAAATCCTACTTCTAAGCAAAATTAATGGTGATGATGCAGATGCC 1794
 QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
 DB 1795 ACATGTACCGAAAGAAATTCAGGTAGCAGCAAGAAATACATGTCGAATGACTATAA 1854
 QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
 DB 1855 CCGTATCTTATCCACTTTTCGATGATGATTTCTGCAGTCC 1896

RESULT 4

PFAMSP814 1729 bp DNA linear INV 29-MAY-2002
 LOCUS P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.

ACCESSION DI3345
 VERSION DI3345.1 GI:391790
 KEYWORDS EGF-like domains; major merozoite surface protein precursor.
 SOURCE Plasmodium falciparum (isolate 814) merozoite, DNA.
 ORGANISM Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 1729)
 Jongsutwittes, S., Tanabe, K. and Kanbara, H.
 Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates
 Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
 93295445

2 (bases 1 to 1729)
 Jongsutwittes, S.
 Direct Submission
 Submitted (05-OCT-1992) Sanchai Jongsutwittes, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (tel:0958-47-2111(ex.3747), Fax:0958-47-6607)
 Submitted (05-OCT-1992) to DDBJ by:

Somchai Jongsutwittes
 Department of Protozoology
 Institute of Tropical Medicine
 Nagasaki University
 12-4 Sakamoto-machi
 Nagasaki 852
 Japan
 Phone: 0958-47-2111x3747
 Fax: 0958-47-6607.

FEATURES

source

1. .1729
 Location/Qualifiers
 /organism="Plasmodium falciparum"
 /db_xref="taxon:5833"
 /chromosome="9"

/dev_stage="merozoite"

gene

1. .1715

/gene="MSP1"

CDS

1. .1715

/gene="MSP1"

/codon_start=3

/product="major merozoite surface protein precursor"

/protein_id="BAA02606.1"

/db_xref="GI:391791"

/translation="PLKLTSEVSIQEDNYANLEKFRVLKIDGLNDLNLHLGKKKLS
 FLSSGLHLITELKWKVKNKNTGNSPENNKKVNEALKSYENFLPEAKVHTVTPPQ
 PDVTPSPLSVRVSGSGSTKEETQPTSGSLLETLQVQVQSONYDEEDSDVVLPIFG
 ESEDNDYLDQVVTGEAISVTMDNLSGFENYDVYIYKPLAGVYRSUKQTEKNIT
 FNLNDILNLSRLKRRKFLDVLDSLMQFKHISSEYIIEISFKLNSQKTEKLLK
 YKIKESVENDIKFAQEGISYEKVLAKYDDLESIKKVIKEKPKFPSPPTTSPSP
 AKTDQKKESKDFLPLFTNLTETLYNNLVNKIDVYLNKAKINDCNVKEDEAHVKTKL
 SPDKAIDDKIDLFKNTDPEATKKLINDPTKMDLGLKLLSTGLVNFNPWTIISKLEIG
 KQDMLNISOHCYVKQCPENSCFRHLDERBECKCLLNYKQEGKCVENPNTCNEN
 NGCCDADATCTEEDSGSRKKITCECTKPDSPFLDFGIFC\$SSNFGLISFLILMLIL
 YSFI"

BASE COUNT 742 a 234 c 246 g 507 t
 ORIGIN

Alignment Scores:

Pred. No.: 1,01e-127 Length: 1729
 Score: 1946.50 Matches: 374
 Percent Similarity: 96.70% Conservative: 7
 Best Local Similarity: 94.92% Mismatches: 8
 Query Match: 85.56% Indels: 5
 DB: 3 Gaps: 1

US-10-057-531A-2 (1-431) x PFAMSP814 (1-1729)

QY 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57
 DB 486 TCCGAGAGATAATCAGCAATATTTAGATCAAGTAGTAGTACGAGAA----- 530
 QY 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValle 77
 DB 531 GCAATATCTGTCAACATGGATAATATCTCTCAGAGATTTGAAATGAATATGATGTATA 590
 QY 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
 DB 591 TATTTAAACCTTTAGCTGGAGTATATAGAGCTTAAACAAATTTGAAACAAATTTGAAACAAATTT 650
 QY 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
 DB 651 ATTACATTTAAATTTAAATTTGAACGATATCTTAAATTCACGCTCTTAAGAAACGAAATAT 710
 QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
 DB 711 TTCTTAGATGTATTAGAAATCTGATTTAATGCAATTTAAACATATATATCTCAAAATGAATAC 770
 QY 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
 DB 771 ATTATTGAAGATTCATTTAAATTTATTTGAATTCAGACAAACAAACACACACTTTTAAACAAAGT 830
 QY 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
 DB 831 TACAATATATAAAGAAATCAGTAGAAATGATATTAATTTGCACAGGAAGGTATAGT 890
 QY 178 TyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValle 197
 DB 891 TATTATGAAAGGTTTTAGCGAAATATAAGGATGATTTAGAAATCAATTAACAAAGTTATC 950
 QY 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
 DB 951 AAAGAAAGAAAGGAGAGAGTTCCCATCATCACCAACCAACACCTCGCTCACCAGCAAA 1010
 QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
 DB 1011 ACAGACGACAAAGAAAGGAAAGTAAAGTCTCTTCCATTTTAAACAAACATTTAGACCTTA 1070
 QY 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
 DB 1071 TACAATAACTTAGTTAATAAAATTGACGATTACTTAATTAACCTAAAGCGCAAGAGTTAAC 1130
 QY 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
 DB 1131 GATTGTAGTGTTCGAAAGAGATGAAGCACATGTTAAAAATAAATAACTTAGTGTATTTAAAA 1190

Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
 Db 1191 GCAATTGATGACAAAATAGATCTTTTAAAAACACTAACGACTTCGAAGCAATTAATAAAA 1250
 Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
 Db 1251 TTGATAAATGATGATACGAAAAAGATGCTTGGCAAAATTTACTTAGTACAGGATTAGTT 1310
 Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGlyLysPheGlnAspMetLeu 337
 Db 1311 CAAATTTTCTTAACTAATAATATCAAAATTAATGAAGGAAATTCACAGATATGTTA 1370
 Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
 Db 1371 AACATTTTCAACACCAATGCGTAAAAAACAATGCTCCAGAAAAATTCGGATGTTTCAGA 1430
 Qy 358 HisLeuAspGluAtrGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
 Db 1431 CATTAGATGAAGAGAGAATGTAATGTTTATTAATTAACAACAAGAGGTGATAAA 1490
 Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
 Db 1491 TGTGTGAAATCCAAATCCCTACTTGTAAACGAAATAATGCTGCGATGTGATGCAGATGCC 1550
 Qy 398 ThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys 417
 Db 1551 ACATGTACCGAAGAAGATTTCAGGTAGCAGCAGCAAGAAAGAAATTCACATGTGAATGACTAAA 1610
 Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
 Db 1611 CCTGATCTTATCCACTTTTCGATGGTATTTTCGACAGTTCC 1652

PFAMSPM20 1729 bp DNA linear INV 29-MAY-2002
 LOCUS P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.
 DEFINITION D13343
 ACCESSION D13343.1 GI:391822
 VERSION EGF-like domains; major merozoite surface protein precursor.
 KEYWORDS Plasmodium falciparum (isolate MAD20) merozoite, DNA.
 SOURCE Plasmodium falciparum
 ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 Tanabe,K., Mackay,M., Goman,M. and Scaife,J.G.
 AUTHORS Allelic dimorphism in a surface antigen gene of the malaria
 TITLE parasite Plasmodium falciparum
 JOURNAL J. Mol. Biol. 195 (2), 273-287 (1987)
 MEDLINE 88011243
 REFERENCE 2 (bases 1 to 1729)
 AUTHORS Jongwutives,S., Tanabe,K. and Kanbara,H..
 TITLE Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates
 JOURNAL Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
 MEDLINE 93295445
 REFERENCE 3 (bases 1 to 1729)
 AUTHORS Jongwutives,S.
 TITLE Direct Submission
 JOURNAL Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology: 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)
 COMMENT Submitted (05-OCT-1992) to DDBJ by:
 Somchai Jongwutives
 Department of Protozoology
 Institute of Tropical Medicine
 Nagasaki University
 12-4 Sakamoto-machi
 Nagasaki 852
 Japan
 Phone: 0958-47-2111x3747
 Fax: 0958-47-6607.

FEATURES
sourceLocation/Qualifiers
1. .1729

/organism="Plasmodium falciparum"

/db_xref="taxon:5833"

/chromosome="9"

/dev_stage="merozoite"

1. .1715

/gene="MSP1"

<1. .1715

/gene="MSP1"

/codon_start=3

/product="major merozoite surface protein precursor"

/protein_id="BAA02604.1"

/db_xref="GI:391823"

/translation="PLKLTSEVSIQTEDYNAVLEKFRALSKIDGKLNLDNLHLGKKLLS
 FLSSGLHLITELKEVTKNNKNTGNSPENNKKVNEALKSYENFLPEAKVTVTVTPPO
 PDVTPSPLSVRVSGSGSKTEETQIPTSGSLLELQOVVOLQNYDEDDSLVLPFG
 ESDNDEYLDQVVTGSAISVTMDNLIISGFENEDYVILKPLAGYVYKSLKQIKENIIT
 FNLNLDILNSRLKKRYFLDVLSDLMQFKHSSNEYIIEEDFKLLNSPKRTLLKS
 YKIKESVENDIKFAQEGISYVEKVLAKYKDDLEIKKVIKEKEKFPSSPPTPPSP
 AKTDEQKESKFLPFTNITLNNLVNKIDDLINLKKLSTGLVONFPNTIISKLEIG
 SDLKAIDDKIDLFKNTDPEAIKKLINDPTKDMGLKLTGLVONFPNTIISKLEIG
 KFDMLNISQHCVKKQCPENSCFRLDERECKLLNKKQSGDKCVENPNTCEN
 NGCDDADATCTEEDSGSSKKITCECTKPSYPLFDGIFGSSNFIQISFLILMLIL
 YSFI"

BASE COUNT 742 a 234 c 246 g 507 t

ORIGIN

Alignment Scores:

Pred. No.: 1,01e-127 Length: 1729
 Score: 1946.50 Matches: 374
 Percent Similarity: 96.70% Conservative: 7
 Best Local Similarity: 94.92% Mismatches: 8
 Query Match: 85.56% Indels: 5
 DB: 3 Gaps: 1

US-10-057-531A-2 (1-431) x PFAMSPM20 (1-1729)

Qy 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGlyArgGlyThrMet 57

Db 486 TCCGAAGATAATGACGAATATTAGATCAAGTACTAGTGGAGAA----- 530

Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValle 77

Db 531 GCAATATCTGTCACAAATGGATAATATCTCTCAGGATTTGAAAATGAATATGATGTTATA 590

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97

Db 591 TATTTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAAACAAATTTGAAAAACATT 650

Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117

Db 651 ATTACATTTAATTTAAATTTGACGATATCTTAAATTCACGCTCTTAAGAAACGAAATAT 710

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137

Db 711 TTCTTAGATGATTAGAAATCTGATTAAATGCAATTTAAACATATATCTCAATGAATAC 770

Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157

Db 771 ATTATTGACATTCATTTAAATTTGCAATTTCAACACAAAAACACACTTTTAAAGT 830

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177

Db 831 TACAATATATAAAGAAATCAGTAGAAATGATATAAATTTGCACAGGAGGTATAAGT 890

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValle 197

Db 891 TATTATGAAAAGGTTTGTAGCGAAATATAAGGATGATTAGAAATCAATTAATAAGTTATC 950

Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217

Db 951 AAAGAAGAAAGGAGAGTTCCTCATCATCACCAACACACCTCCCTCACCAGCAAAA 1010

BASE COUNT 2376 a /note="pot.glycosylation site"

ORIGIN 685 c 702 g 1629 t

Alignment Scores:

Pred. No.: 3 33e-127 Length: 5392
Score: 1946.50 Matches: 374
Percent Similarity: 96.70% Conservative: 7
Best Local Similarity: 94.92% Mismatches: 8
Query Match: 85.56% Indels: 5
DB: 3 Gaps: 1

US-10-057-531a-2 (1-431) x PFP190G1 (1-5392)

Qy 38 ThrAspAspAspLysAlaMetAlaAspLleGlySerLleGluGlyArgGlyThrMet 57
Db 4012 TCCGAAGATAATGACCAATATTAGATCAAGTAGTAACCTGGAGAA-----4056

Qy 58 AlalleSerValThrMetAspAsnLleLeuSerGlyPheGluAsnGluTyrAspValle 77
Db 4057 GCAATATCTGTACAAATGGATAATATCTCTCAGGATTGGAATGAATGATGTTATA 4116

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnLleGluLysAsnLle 97
Db 4117 TATTTAAACCTTTAGCTGGAGTATATAGAACCTTAAANAACAATTTGAANAACAATT 4176

Qy 98 PheThrPheAsnLeuAsnLeuAsnAspLleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 4177 ATTACATTTAATTTAAATTTGAACGATATCTTAATTTCCAGCTCTTAAGAAACGAAATAT 4236

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisLleSerSerAsnGluTyr 137
Db 4237 TCTTTAGATGATTAGATCTGATTGATTAATGCAATTTAAACATATATCTCAATGAATAC 4296

Qy 138 IleLleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
Db 4297 ATTATTGAAGATTCATTTAAATTTAGTATTCAGAACCAAAAAACACACTTTTAAANAAGT 4356

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspLleLysPheAlaGlnGluGlyLysSer 177
Db 4357 TACAATATATAAAGAATCAGTAGAATAATGATATTAATTTGCACAGGAAGGTATAAGT 4416

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValle 197
Db 4417 TATTATGAAGAAGTTTACGCAATATAGGATGATTAGATCAATTTAAANAAGTTATC 4476

Qy 198 LysGluGluLysGluLysPheProSerProProThrThrProProSerProAlaLys 217
Db 4477 AAAGAAGAAAGGAGAGAGTCCCATCATCACCAACCAACACCTCCGTCACCAAGCAAAA 4536

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnLleGluThrLeu 237
Db 4537 ACAGACGAACAAAGAAGAAAGTAAATTCCTTCATTTTAAACAAACATTGACAGCCTTA 4596

Qy 238 TyrAsnAsnLeuValAsnLysLysAspTyrIleLeuAsnLeuLysAlaLysLysAsn 257
Db 4597 TACAATACTTAGTTAATAAATTGACCGATTACTTAATTAATTAACGCAAGCAATTAAC 4656

Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysLleThrLysLeuSerAspLeuLys 277
Db 4657 GATTGTAATGTTGAAAGAGATGAGCACATGTTTAAATAACTAACTTAGTATTGATTAATA 4716

Qy 278 AlalleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaLleLysLys 297
Db 4717 GCAATTTGATGACAAATAGATCTTTTAAAAACACATAACGACTTCGGAAGCAATTAATA 4776

Qy 298 LeuLleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
Db 4777 TTGATAAATGATGATACCAAAAAAGATATGCTTGCAAAATTACTTAGTACAGGATTAGT 4836

Qy 318 GlnAsnPheProAsnThrIleLleSerLysLeuLleGluGlyLysPheGlnAspMetLeu 337
Db 4837 CAAATTTCTCTATACAAATATATCAAAATTAATTAAGGAANAATTCGAAGATATGTTA 4896

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 4897 AACATTTTCCAAACACCAATGCGTAAAAAACAATGTCAGAAAAATTCGTGATGTTTTCAGA 4956

Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
Db 4957 CATTTAGATGAAAGAGAAGAATGTAATGTTTATTTAAATACAAACAAGAGGTGATAAA 5016

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
Db 5017 TGTGTTGAAATCCAAATCCTACTTGTACGAAAAATATGTTGGATGTGATCAGATGCC 5076

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysLleThrCysGluCysThrLys 417
Db 5077 ACATGTCACGAAGAAGATTACAGGTAGCAGCAGAGAAAAATACACATGTGAATGTACTAAA 5136

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
Db 5137 CCGTATCTTATCCACTTTTTCGATGTTATTTTCTCAGTTCC 5178

RESULT 7

PFAMMSAA 5754 bp mRNA linear INV 14-MAR-1994
LOCUS P.falciiparum major merozoite surface antigen (PMMSA) mRNA, complete
DEFINITION cds, isolate FC27.
ACCESSION M19143.1 GI:160412
VERSION antigen; glycoprotein.
KEYWORDS P.falciiparum (isolate FC27 from Papua New Guinea), cDNA to mRNA,
SOURCE clone Ag75, gl.1, gl26, pepG3.3.
ORGANISM Plasmodium falciiparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 5754)
AUTHORS Peterson,M.G., Coppel,R.L., McIntyre,P., Langford,C.J., Woodrow,G.,
Brown,G.V., Anders,R.F. and Kemp,D.J.
TITLE Variation in the precursor to the major merozoite surface antigens
of Plasmodium falciiparum
JOURNAL Mol. Biochem. Parasitol. 27 (2-3), 291-301 (1988)
MEDLINE 88142999
PUBMED 2449612

FEATURES
Location/Qualifiers
1..5754
/organism="Plasmodium falciiparum"
/isolate="FC27 (Papua, New Guinea)"
/db_xref="taxon:5833"
1..5754
/gene="PMMSA"
<1..5754
/gene="PMMSA"
121..5226
/gene="PMMSA"
/note="precursor"
/codon_start=1
/product="major merozoite surface antigen"
/protein_id="AAA29653.1"
/db_xref="GI:160413"
/translation="MKIIFLCSEFLFIINTQCVTHESYQELVKKLEALEDAVLTYGS
LFQKRMVLNGETSGTAVTTSTPGSGSVTSGSVASVASVSGSGSGSGSGSGS
RTPNPSNDSDNTKTVDLKHRYONYLFTIKELYPFLDLTNHMLTLSKNVDFKY
LIDGYEINELLYKLFYDILRAKLANDACANSYQIPFNKLIRANELDVLKLVFGY
RKPLDINKDVGKMEYIKNKTITIANINELTIGSKKTIIDKNADNEEGKKLYAQ
YNFYINKQLOEAMNLSVLEKRIIDILKNENIKKLEDDIKITKDEADVPTGSKPNP
LPENKKEVEGHEEKIKEIARTIKFNDSIDTDPLELYLREKNKVDVTPKSDPT
KSPQIPKVPNGVYPLPTDIHNSLAADNKSQGLMNPDKKEINEXIITDNKE
RKFIINIKKQIDLEEKINHTKEQNKLLDEYKSKDYFELLEKEVEMKFNNFQK
DVVDKIFESARTYNYVQRYNKKNSSNNSSNVYVQKLLKHALSYLEDLSLRKISEKDF
NHYYTLKGLADIKLITEIKSENKILEKNFKLTHSANALESVDIVKLVQKVL
LIRKIEDLRKIELFKNAQLKDSIHVPNIYKPNKPEPYLVILVKEDVLEKEIPKV
KMLKKEQAVLSSTIOPAASETTEGHSHTLSQSGETEVTEETVEETVIGHT
TVTTLTPKEESAPKEVKVVENSTEHSNDNSOALTKTVYLKLUDEFLTYSYCHKYI
LYSNSMDOKLLEVYNLTPREENELKSCDPLDLTFNIONNIPAMYLYSDSNIDLOHL
FFELIQEMYIYHLKKEENHKKLEEQKQITGSTSTSSPGNTVTVTAQASHNSQ

NQOSNASSNTNTQNGVAVSSGPVAVVEESHDLPTVLSTISNDLKGIVSLNLLNGKTKVVPN
 LFTISPTMEKFFYENILKUNNDTTFNDDIKQFVKSNSKVITGLTETQKNAUNDLTKIKLUD
 TIQSLSDYKLLKLDORFNKELGQDKMQIKKLTILKLEQLESKLNPNPNHLYQ
 NFSVPFNKKKEAEIAGTENTLTKLLHKGVLGVYNGESPSPKLTSEVSYTQSTEDN
 VANLEKFRALKKGLNDLNLHGLKKLSFLSSGLHHLITELKEVITKNKNTYTCNSPE
 NKKYNEALKSVENELPEAKYTVTVTPPOPDVTPSPLSRVSSSGSTSEETQIPIPSG
 SLUTEQQVYQLQNYDESDSLVLPFI GESEDNDEYLDQVTVTGEAISVYTMNDLSLQ
 ENEDYVILPKLAGVYRDKKQKTEKNIITFNGLNDLNLRLKRRKFFLDVLSLDMQ
 FKHSINSEYIIEDSFKLNSQKNTLLKSVKYIKEVENDIKFAQEGISYVEYKVLAKY
 KDOLESIKVIEKEEKFPSPPTTPSPAKTDEQKESKFLPFLNIETLYNLYNKY
 IDDYILNLKAKINDCNVEKDEAHVKITKLSLKAIDDKIDLEKNTNDFEAIKKLINDD
 TKKMDGLKLSLGLVQNFNTIISKLEKGFQDMLNLSHQCVKSGKQSPNSGCGFRHLD
 ERECKCLNLYKQEGDKVQENPNTNENNGGDADATCEEDSSGRSKITCECTKPK
 DSVPLFDGIFCNSNPLGFSFLILMLIYSPF"
 121. 177

```

BASE COUNT
ORIGIN
2539 a . 709 c 708 g 1798 t
218 bp upstream of HindIII site.
/product="major merozoite surface antigen"
/ gene= fmsA

```

Alignment Scores:		
Pred. NO.:	3.56e-127	Length:
Score:	1946.50	Matches:
Percent Similarity:	96.70%	Conservative:
Best Local Similarity:	94.92%	Mismatches:
Query Match:	95.56%	Indels:
DB:	3	Gaps:
	5	
		5754

US-10-057-531A-2 (1-431) x PFAMMSAA (1-5754)

Qy	36	ThrAspAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet	57
Db	3997	TCCGAAAGATAATCAGCAATATTTAGCATCAAGTAGTAAGTCAGAGAA-----	4041
Qy	58	AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle	77
Db	4042	GCAAATATCTGCACAAATGGATTAATCTCTCAGGATTTGAAATGAATGATGATGTTATA	4101
Qy	78	TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle	97
Db	4102	TATTTAAAAACCTTTAGCTGGAGTATATAGAAGCTTAAAAAACAACAAATGAAAAAACATTT	4161
Qy	98	PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr	117
Db	4162	ATTACATTTTAAATTTAAATTTGAACCATATCTTAAATTCACGCTTTAAGAACGAAAAATAT	4221
Qy	118	PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr	137
Db	4222	TTCTTAGATGTATTAGAAATCTGATTTAATGCAATTTAAACATATATATCTCAATGGAATAC	4281
Qy	138	IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer	157
Db	4282	ATTATTGAAGATTCAATTTAAATTTATTGGAATTCAGAACAAAAAACAACACTTTTAAAAAGT	4341
Qy	158	TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer	177
Db	4342	TACCAATATATAAAGAATCAGTAGAAATGATATAATTAATTTGCACGAGGAAGGTATAAGT	4401
Qy	178	TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle	197
Db	4402	TATTATGAAAGGTTTTACGCGAAATATAAGGATGATTTAGAAATCAATTAATAAAGATTTATC	4461
Qy	198	LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys	217
Db	4462	AAAGAGAAAAAGGAGAAGTTCCCATCATCACACCACCAACAACACCTCCGCTCACCAACAAAA	4521
Qy	218	ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu	237
Db	4522	ACAGACGAACAAGAAAGAAAGTAAGTTCCCTTCATTTTTTAAACAACAACATTTGACACCTTA	4581
Qy	238	TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn	257

```
/codon_start=1
/product="merozoite surface protein 1, 42 kDa C-terminal
/region"
/protein_id="AAA62219.1"
/db_xref="GI:664858"
/translation="AISTVMDNISLGFENYDVLYLKLACVYSLKKQIEKNIFTFN
LNLNLSRLKKRKYFLDVLDSLDFKHLSSNEYIEDSFKLNSQKNTLLKSYK
YIKESVENDIKFAQGGISYEKVLAKYKDDLESIKKVIKEERKFPSPPTPPSPAK
TDEQKESKFLPFTNLETETLYNNLVNKIDDDYILNLKAKINDCNVDEARHVIKPLSD
LKAIIDDKIDLFKNPYDEAIKKLINDDTTKKMLGKLLSTGLVQNPNTIISKLEGRF
ODMLNIHQHCVKQCPQNSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNG
GCDADAKCTEEDSNGKKITCECTKPDSPFLFDGIFCS"
BASE COUNT 485 a 147 c 161 g 326 t
ORIGIN
```

```
Alignment Scores:
Pred. No.: 1,32e-127 Length: 1119
Score: 1942.00 Matches: 369
Percent Similarity: 99.46% Conservative: 2
Best Local Similarity: 98.93% Mismatches: 2
Query Match: 85.36% Indels: 0
Db: 3 Gaps: 0
```

US-10-057-531A-2 (1-431) x PFU20728 (1-1119)

```
Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
Db 1 GCAATATCTGCACAAATGGATAATATCTCTCAGGATTGAAATGAAATGATGTTATA 60
Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
Db 61 TATTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAAACAATTTGAAAAACAAT 120
Qy 98 PheThrPheAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 121 TTTACATTAATTAATTTGAACGATATCTTAAATTCACGCTCTTAAGAAACGAAATAT 180
Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
Db 191 TTCCTAGATGATTAGAACTGATTAAATGCAATTTAAACATATATCTCAATCAATAC 240
Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
Db 241 ATTATTGAAGATTTCATTTAAATATTGTAATTCAGACACAAAAACACACTTTTAAAGT 300
Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
Db 301 TACAAATATATAAAGAAATCAGTAGAAATGATATTAATTTGCACAGGAAGGTATAAGT 360
Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle 197
Db 361 TATTATGAAAAGGTTTATAGCGAAATATAAGGATGATTAGAAATCAATTAATAAGTTATC 420
Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
Db 421 AAAGAGAAAAGGAGAGAGTTCCCATCATCATCACCAACACACCTCCGTCACAGCAAAA 480
Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
Db 481 ACAGACGACAAAAGAGAGAAAGTAAAGTTCTCTCATTTTAAACAAACATTGACACCTTA 540
Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrIleLeuLeuAsnLeuLysAlaLysIleAsn 257
Db 541 TACAAATACCTTAGTTAATAAATTTAGCATTTACTTAATTAATTAATTAATTAATTAAC 600
Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
Db 601 GATTGTAATGTTGAAAAGAGATGAAGCAGCATCTTAAATACTAACTAGCTAGTTTAAAA 660
Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
Db 661 GCAATTGATGACAAAATAGATCTTTTTTAAAAACCCCTTACGACTTCGAAGCAATTAATAA 720
```

```
Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
Db 721 TTGATAAATGATGATACGAAAAAAGATATGCTTGCATAATTACTTAGTACAGATTAGTT 780
Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluLysPheGlnAspMetLeu 337
Db 781 CAAAATTTTCTTAATACCAATTAATCAAAATTAATTAAGGAAAAATTCACAGATATGTTA 840
Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 841 AACATTTCCACACACCAATGCGTAAAAAACAATGCTCCACAAAATTCGGATGTTTTCAGA 900
Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluLysAspLys 377
Db 901 CATTTAGATGAAGAGAGAAATGTAATGTTTATTAAATTCACAAACAAGAGTGATATAA 960
Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
Db 961 TGTGTTGAAAATCCAAATCCTACTTGTAAACGAAAAATTAATGTTGGATGTCAGATGCC 1020
Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
Db 1021 AAATGTACCGAAGAGATTTCAGTAGCAACGGAAGAAATAACACATGTGAATGTACTATA 1080
Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
Db 1081 CCGTATCTTATCCACTTTTCGATGTTATTTCTGCAGT 1119
RESULT 9
PFAMSP841 1726 bp DNA linear INV 29-MAY-2002
LOCUS P. falciparum DNA for the precursor to the major merozoite surface
DEFINITION proteins, C-terminal.
ACCESSION D13351
VERSION D13351.1 GI:391810
KEYWORDS EGF-like domains; major merozoite surface protein precursor.
SOURCE Plasmodium falciparum (isolate 841) merozoite, DNA.
ORGANISM Plasmodium falciparum
Eukaryotes; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
REFERENCE 1 (bases 1 to 1726)
AUTHORS Jongwutiwes, S., Tanabe, K. and Kanbara, H.
TITLE Sequence conservation in the C-terminal part of the precursor to
the major merozoite surface proteins (MSP) of Plasmodium
falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
MEDLINE 93295445
REFERENCE 2 (bases 1 to 1726)
AUTHORS Jongwutiwes, S.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1992) Somchai Jongwutiwes, Institute of Tropical
Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4
Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747),
Fax:0958-47-6607)
COMMENT Submitted (05-OCT-1992) to DBJ by:
Somchai Jongwutiwes
Department of Protozoology
Institute of Tropical Medicine
Nagasaki University
12-4 Sakamoto-machi
Nagasaki 852
Japan
Phone: 0958-47-2111x3747
Fax: 0958-47-6607.
FEATURES
source
1. .1726
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="9"
/dev_stage="merozoite"
1. .1712
/gene="MSP1"
<1. .1712
/gene="MSP1"
gene
CDS
```

```

/codon_start=3
/product="major merozoite surface protein precursor"
/protein_id="BAA02612.1"
/db_xref="GI:391811"
/translation="PLKVLSEVSIQEDNYANLEKFRALSKIDGKLNLDNLHLGKKKLS
FLSGHLHLITELKEVYIKNNKNTGNSPENNKKVNEALKSYENFLPEAKVTVTVPQ
PDVTPSPISVRVSGSGSTKEETQIPTSGSLTDELQVQVQLONYDEEDDSLWLPFG
ESEDNDYLDQVVTGEAIVSTMNDILSGFENEYDVIYIKPLAGVYRSIKKQIEKNIFT
FNLNLDLNSRLKKRKYFLDVLSDLMQFKHISSEYVIEDSFLLANSEOKNTLLKS
YKYLKESVENDIKFAQEGISYKVLAKYKDDLESIKKVIKEEREPSPPTPSPA
KTDGKKSKSLPFLUTNLETLYNNLVNKIDDLINLAKINDCNVKEDEARVKKTKLS
DLKADDDIKVKNPDNFEAIKKLINDDKKMLKLLSTGLVQNFNTIISKLEKG
FQMLNLSHQCVKQCPENSGCFRHLDERECKLNYKQEGDKCVENPNPTCENN
GGCDAATCTBEDSGSRKKITCECTKPDSPYPLFDGIFCFSNLFGLISFLILMLLY
SFI"

```

BASE COUNT 739 a 234 c 245 g 508 t

Alignment Scores:

Pred. No.: 4, 67e-127 Length: 1726
 Score: 1937.00 Matches: 374
 Percent Similarity: 96.70% Conservative: 7
 Best Local Similarity: 94.92% Mismatches: 7
 Query Match: 85.14% Indels: 6
 DB: 3 Gaps: 2

US-10-057-531A-2 (1-431) x PFAMSP841 (1-1726)

QY	38	ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet	57
DB	486	TCCGAAGATAATGACGAATATTATGATCAAGTAGTAAGTGGAGAA-----	530
QY	58	AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle	77
DB	531	GCAATATCTGCACATGATATATATCTCTCAGGATTTGAAATGAATATGATGTATA	590
QY	78	TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle	97
DB	591	TATTTAAACCTTTAGCTGAGTATATAGAGCTTAAAAAACAATAATGAAAAACAAT	650
QY	98	PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr	117
DB	651	TTTACATTTAAATTTGAACGATATCTTAAATTCACGCTTTAAGAAACGAAATAT	710
QY	118	PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr	137
DB	711	TTCTTAGATGATATAGAAATCTGATTTAATGCAATTTAAACATATATCTCAATGAATAC	770
QY	138	IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer	157
DB	771	ATTATTGAAGATTTCATTTAAATTTATGAAATTCAGAACAAAAACACACTTTTAA	830
QY	158	TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer	177
DB	831	TACAAATATATAAAGAATCAGTAGAAATGATATTAATTTGCACAGGAAGTATAGT	890
QY	178	TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle	197
DB	891	TATTATGAAAGGTTTATAGGAAATATAAGGATGATTTAGATCAATTAATAAAGTTATC	950
QY	198	LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys	217
DB	951	AAAGAAGAAAGAG---TTCCCATCATCACCACCAACACACCTCCGTCACCCAGCAAA	1007
QY	218	ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu	237
DB	1008	ACAGACACACAAAGAAGAAAGTAAGTTCTTCCATTTTAAACAACATTTGACACCTTA	1067
QY	238	TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn	257
DB	1068	TACAATAACTTAGTTAATAAATGACGATTACTTAATTAACCTTAAGGCAAGATTAAAC	1127
QY	258	AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys	277

DB	1128	GATTGTAATGTTGAAAAAGATGAACACATGTTAAAAATAACTTAACCTTAGTGATTTAAAA	1187
QY	278	AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys	297
DB	1188	GCAATTTGATGACAAAAATAGATCTTTTAAAAACACTTAACGACTTCGAAGCAATTA	1247
QY	298	LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal	317
DB	1248	TTGATAAATGATGATACGAAAAAAGATATGCTTGGCAAAATTACTTAGTACAGGATTAGTT	1307
QY	318	GlnAsnPheProAsnThrIleLeSerLysLeuIleGluLysPheGlnAspMetLeu	337
DB	1308	CAAAATTTTCTTAATAACAATAATATCAAAATTAATTTGAAGGAAAAATTTCCAAGATATGTTA	1367
QY	338	AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg	357
DB	1368	AACATTTTCACACACCAATGCGTAAAAAACAATGTCAGAAAAATTTCTGGATGTTTCAGA	1427
QY	358	HisLeuAspGluArgGluGluCysLysCysLysLeuLeuAsnTyrLysGlnGluGlyAspLys	377
DB	1428	CATTTAGATGAAGACAGAATGTAATGTTTATTAATTAACAACAAGAGGTGATAAA	1487
QY	378	CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla	397
DB	1488	TGTGTTGAAAAATCCAAATCTACTTGTAAACGAAAAATAATGGTGGATGTCAGATGCC	1547
QY	398	ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys	417
DB	1548	ACATGTACCGAAGAGATTGAGTAGCAGCAGCAAGAAAAATACATGTGAATGTACTATAA	1607
QY	418	ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431	
DB	1608	CCTGATTCATTACCACTTTTCGATGCTATTTCTGCAGTCC 1649	
RESULT	10		
PF0195			
LOCUS			
DEFINITION		5541 bp DNA linear INV 12-SEP-1993	
		P.falciparum (CAMP strain) gp195 gene for major merozoite surface antigen.	
ACCESSION		X03831 X06887	
VERSION		X03831.1 GI:9894	
KEYWORDS		glycoprotein; surface antigen; tandem repeat.	
SOURCE		Plasmodium falciparum.	
ORGANISM		Plasmodium falciparum	
REFERENCE		Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
AUTHORS		1 (bases 1 to 3593)	
TITLE		Weber, J.L., Leininger, W.M. and Lyon, J.A.	
JOURNAL		Variation in the gene encoding a major merozoite surface antigen of the human malaria parasite Plasmodium falciparum	
MEDLINE		Nucleic Acids Res. 14 (8), 3311-3323 (1986)	
PUBMED		86205236	
REFERENCE		2 (bases 3581 to 5396)	
AUTHORS		Weber, J.L.	
TITLE		Direct Submission	
JOURNAL		Submitted (12-MAR-1987)	
REFERENCE		3 (bases 3594 to 5541)	
AUTHORS		Weber, J.L., Sim, B.K., Lyon, J.A. and Wolff, R.	
TITLE		Merozoite surface protein sequence from the Camp strain of the human malaria parasite Plasmodium falciparum	
JOURNAL		Nucleic Acids Res. 16 (3), 1206 (1988)	
MEDLINE		88143999	
PUBMED		3278296	
COMMENT		Data kindly reviewed (12-MAR-1987) by Weber J.L.	
FEATURES		Location/Qualifiers	
source		1..5541	
		/organism="Plasmodium falciparum"	
		/strain="CAMP (Malaysia)"	
		/db_xref="taxon:5833"	
		283..5463	
		/note="pot. major merozoite surface antigen"	
		/codon_start=1	

CDS

(MSP-1) in field-derived Plasmodium falciparum
Unpublished
2 (bases 1 to 1119)
Shi, Y.-P.
Direct Submission
Submitted (07-FEB-1995) Ya-Ping Shi, DPD/CDC 4770, Molecular
vaccine Section, Bldg 22, Rm. 8F12, Buford Highway, Chamblee, GA,
30341 USA

Location/Qualifiers
1. .1119
/organism="Plasmodium falciparum"
/isolate="infected human blood sample from Papua New
Guinea"
/db_xref="taxon:5833"
/clone="K46B"
1. .1119
/gene="MSP-1"
/gene="MSP-1"
/codon_start=1
/product="merozoite surface protein 1, 42 kDa C-terminal
region"
/protein_id="AA62218.1"
/db_xref="GI:664856"
/translation="AISVTMDNILSGFENEDYIYLRPLAGVYRSKKQIKNIPTFN
LNLIDILSRKRRKYFLDVEISLMQFKHSSNEYIIEDESKLINSQKNTLLKSYK
YIKESVNDIKFAOEGISYVEKVLAKYKDDLESIKVKEKEKFPSSPPPTPPSPAK
TDEQKKESKFLPFTNLETLYNNLVNKIDHYLNLKAKINDCNVEKDHVKIKLSD
LKAIDDKLFPNYPDEAIAKKLINDTKDKMLGKLLSTGLVQFPNTIISKLEGRF
QMDNLISQVCVKQPCQNSCFRHLDERECKLLNLYKQEGDKCVENPNPTCNEEN
GCDADAKTEDSGSKKTKCTCKPDSYPLFDGIFCS"

BASE COUNT 485 a 148 c 160 g 326 t
ORIGIN

Alignment Scores:
Pred. No.: 4,09e-127 Length: 1119
Score: 1935.00 Matches: 368
Percent Similarity: 99.20% Conservative: 2
Best Local Similarity: 98.66% Mismatches: 3
Query Match: 85.05% Indels: 0
DB: 3 Gaps: 0

US-10-057-531A-2 (1-431) x PFU20727 (1-1119)

Qy 58 AlalleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
|||||
Db 1 GCANATATCTGCACATGATATATCTCTCAGGATTTGAAATGAATATGATGTATA 60
|||||

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
|||||
Db 61 TATTTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAACAAATTTGAAAAACATT 120
|||||

Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
|||||
Db 121 TTTACATTTAAATTTAAATTTGAACGATATCTTTAAATTCACGCTTAAAGAAACGAAATAT 180
|||||

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
|||||
Db 181 TTTCTAGATGATATAGAAATCTGATTTAAATGCAATTTAAACATATATCTTCAATGAATAC 240
|||||

Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
|||||
Db 241 ATTATTGAAGATTCATTTAAATTTAATTTGAATTCAGAACAAAAACACACTTTTAAAAAGT 300
|||||

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluLysIleSer 177
|||||
Db 301 TACAAATATATAAAGAAATCAGTAGAAATGATATTAATTTTCACAGGAAGGTATAAGT 360
|||||

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuSerIleLysLysValIle 197
|||||
Db 361 TATATGAAAGGTTTTAGCGAATATAAGGATGATTTAGATCAATTTAAAAAAGTTATC 420
|||||

Qy 198 LysGluGluLysLysPheProSerProProThrThrProProSerProAlaLys 217
|||||

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

Db 421 AAAGAAGAAAGGAGAAGTTCCCATCATCACCAACCAACACCTCCGTCACCAAGCAAAA 480
|||||

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
|||||

Db 481 ACAGACGAACAAAGAAAGAAAGTAAAGTTCTTCCATTTTAAACAAACATTCAGACCTTA 540
|||||

Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
|||||

Db 541 TACATAACTTAGTTAATAAATTTGACCATTACTTAATTAACTTAAAGCAAGATTAAC 600
|||||

Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
|||||

Db 601 GATTGTAAATGTTGAAAAAGATGAAGCACATGTTAAAAATACTTAACCTTAGTGAATTTAAAA 660
|||||

Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaLysLys 297
|||||

Db 661 GCAATTTGATGACAAAATAGATCTTTTAAAAACCTTTACGACTTCGAAGCAATTTAAAAAA 720
|||||

Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
|||||

Db 721 TTGATAAATGATGATACGAAAAAGATATGCTTGGCAAAATTACTTAACAGGATTAAGTT 780
|||||

Qy 318 GlnAsnPheProAsnThrIleLeuSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
|||||

Db 781 CAAAATTTCTTAATACAAATAATCAAAATTAATTTGAAGAAAAATTTCCAAGATATGTTA 840
|||||

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
|||||

Db 841 ACATTTTCAACACCAATGCGCTTAAAAACCAATGTTCCAAAAATTTCTGGATGTTTCAGA 900
|||||

Qy 358 HisLeuAspGluArgGluGlyCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
|||||

Db 901 CATTAGATGAAGAGAGAAGTAATGTAATGTTTATTAATTAACAAACAGAGGTGATAA 960
|||||

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlnGlyCysAspAlaAspAla 397
|||||

Db 961 TGTGTTGAAATCCCAATCTTACTTTGTAACGAAAAATAATGTTGGATGTCAGATGCC 1020
|||||

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
|||||

Db 1021 AAATGTACCGAAGAGATTCAGTACGACGGAAGAAATACATGTGAATGACTATAA 1080
|||||

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
|||||

Db 1081 CCTGATTTTATCCACTTTTCGATGCTATTTCTGCACT 1119
|||||

RESULT 12
PFMEZSA1A
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

PFMEZSA1A 1956 bp DNA linear INV 22-SEP-1995
P.falciparum (RO-33) gene for merozoite surface antigen 1.
235326 M35727
235326.1 GI:535247
merozoite surface antigen 1 gene.
Plasmodium falciparum.
Plasmodium falciparum.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
Tolte, R., Bujard, H. and Cooper, J.A.
Plasmodium falciparum: variations within the C-terminal region of
merozoite surface antigen-1
Exp. Parasitol. 81 (1), 47-54 (1995)
95354793
7628566
2 (bases 1 to 1956)
Tolte, R.
Direct Submission
Submitted (18-JUL-1994) Ralf Tolte, Abt. Prof. H. Bujard, Zentrum
fur Molekulare Biologie (ZMBH), Univ. Heidelberg, Im Neuenheimer
Feld 282, Heidelberg, 69120, Germany
Location/Qualifiers
1. .1956
/organism="Plasmodium falciparum"

```

/Isolate="RO-33"
/db_xref="taxon:5833"
/chromosome="9"
/dev_stage="blood stage"
1..1956
/gene="gpi90, MSAL, MSP1, PMMSA"
<1..1956
/gene="gpi90, MSAL, MSP1, PMMSA"
/codon_start=1
/product="merozoite surface antigen 1 gene, 3+-end"
/protein_id="CAA8455.1"
/db_xref="GI:535248"
/db_xref="SWISS-PROT:PI9598"
/translocation="FNKKELGQDKMOIKLTLKLEQSKLSLNNPHNVLQNFV
FNKKAEAEAEENTENTLILKHYKGLVYNGESSPLKTLSEYIOTEDNYANLE
KFRVLKIDGLNDLHLGKKKFLSSGLHLITELKEVINKNKNTGNSPSENKKV
NEALKSYENFLPAKVTVTVPQDPSPPLSVRVSGSGSPKEETOIPTSGSLITE
LQOVVOLQNYDEEDDSLVLPPIGESEDNDEYLDQVVTGEAISVMDNLSGSENYD
VITKPLAGVYRSKKOIKENIETNLNLDILNSLRKKRYFLDLESDLMQFKHIS
SNBYIIEFSFKLNSOKNTLKSXYIKESVENDIKFAQEGISYERVLAKYKDDLE
SIKKVIEKESEFPSPPTPPSPAKTDEQKESKFLPFLNIETLNNLVNKIDYLI
NLKAKINDCNVEKDEAHVKITKLSDLKATDDKIDLPKNPVDFAIKKLNDPDKOML
KLLSTGLVONPPNTIISKLEGFQDMNLNISOHCVKKOCPONSCGFRHLDRRECK
CLLYNQEGDKQVENPNTCNENNGCDADAKCTEEDSGSNGKKITCECTKPDYPLF
DGIFCSNPLGLSFLILMLILYSFI"
BASE COUNT      853 a      264 c      265 g      574 t.
ORIGIN
Alignment Scores:
Pred. No.:      7,35e-127      Length:      1956
Score:          1935.00      Matches:      372
Percent Similarity: 96.70%      Conservative: 9
Best Local Similarity: 94.42%      Mismatches: 7
Query Match:      85.05%      Indels:      6
DB:              3      Gaps:      2
US-10-057-531A-2 (1-431) x PFMEZSALA (1-1956)

```

```

Qy 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyA-gGlyThrMet 57
Db 730 TCCGAAGATAATAGCAATATTAGATCAAGTAGTAGTGAGAA-----774
Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValle 77
Db 775 GCAATATCTGTCAATAGGATAATATCTCTCAGGATTTGAAATGAATATGATGTTATA 834
Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnile 97
Db 835 TATTTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAACAAATTTGAAAAACATT 894
Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 895 TTTACATTTAAATTTAAATTTGACGATATCTTAAATTCAGCTCTTAAGAAACGAAATAT 954
Qy 118 PheLeuAspValLeuGluSerAspLeuMetClnPheLysHisIleSerSerAsnGluTyr 137
Db 955 TTTCTAGATGATTAGAAATCTGATTTAATGCAATTTAAACATATATCTCAATGAATAC 1014
Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
Db 1015 ATTATTCAAGATTCATTTAAATATTGCAATTCAGACAAAAAACACACITTTAAAAAGT 1074
Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
Db 1075 TACAAATATATAAAGAAATCAGTAGAAAAATGATATTAATTTGCACAGGAGGTATAAGT 1134
Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValle 197
Db 1135 TATTATGAAAGGTTTATTAGCAAAATATAAGATGATTAGATCAATTAATAAAGATTATC 1194
Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
Db 1195 AAGAAGAAAAAGGAG---TTCCCATCATCATCCACCAACACACACCTCGCTCACCAGCAAA 1251

```

```

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrIleu 237
Db 1252 ACAGACCAACAAAGAGAGAAAGTAAGTCTCTCCATTTTAAACAACATTTGACACCTTA 1311
Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
Db 1312 TACAATAACTAGTTAATAAATTCAGGATTACTTAATTAACTTAAAGGCAACAGATTAAAC 1371
Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
Db 1372 GATTGTAATGTTGAAAAGATGAAGCACATGTTAAATAAATACTAAACCTAGTATTAAAA 1431
Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
Db 1432 GCAATTCATGACAAAATAGATCTTTTAAAAACCCCTACGACTTCGAAGCAATTAATAAA 1491
Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
Db 1492 TTGATAATGATGATGACGAAAAAGATATGCTTGGCAATTTACTTAGTACAGGATTAGTT 1551
Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
Db 1552 CAAAATTTTCTATATACATAATATCAAAATTAATTAAGGAAAAATTCGAAGATATGTTA 1611
Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 1612 AACATTTCAACAACACCAATGCGTAAAAAAACAATGTCCACAAAATTTCTGGATGTTTCCA 1671
Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
Db 1672 CATTTAGATGAAAGAGAGAAATGTAATGTTTAAATTAATAAACAAGAAAGGTGATAAA 1731
Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
Db 1732 TGTGTTGAAAATCCAAATCTACTTGTACGAAAATTAATGGTGGATGTGATGCAGATGCC 1791
Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
Db 1792 AAATGTACCGAAGAAAGATTTCAGGTAGCAGGAAAGAAATACATGTGAATGTACTAAA 1851
Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
Db 1852 CCGATTTCTTATCCACTTTTCGATGGTATTTCCTGCGAGTTCC 1893
RESULT 13
PFU20732
LOCUS PFU20732 1119 bp DNA linear INV 14-FEB-1995
DEFINITION Plasmodium falciparum clone B17A merozoite surface protein 1 (MSP-1) gene, 42 kDa C-terminal region, partial cds.
ACCESSION U20732
VERSION U20732
KEYWORDS U20732.1 GI:664865
SOURCE malarial parasite.
ORGANISM Plasmodium falciparum
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS Shi,Y.-P., Alpers,M.P., Povoa,M.M., Nahlen,B.L., Oloo,A.G. and Lal,A.A.
TITLE Sequence of the C-terminal region of merozoite surface protein-1 (MSP-1) in field-derived Plasmodium falciparum
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1119)
AUTHORS Shi,Y.-P.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-1995) Ya-Ping Shi, DPD/CDC 4770, Molecular Vaccine Section, Bldg 22, Rm. 8F12, Buford Highway, Chamblee, GA, 30341 USA
FEATURES
source Location/Qualifiers
1..1119 /organism="Plasmodium falciparum"
/isolate="infected human blood sample from Papua New Guinea"

```


Db	601	GATTGTAATCTGTGAAAAAATGAAGCACACATGTTAAAAATACTAAACTTAGTCATTAAAA	660
Qy	278	AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys	297
Db	661	GCAATTGATCACAAAATAGATCTTTTAAAAACACTAAGCACTTCGAGCAATTAATAAAA	720
Qy	298	LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal	317
Db	721	TTGTATAATGATGATACGAAAAAGATATGCTTGCGCAAAATTACTTAGTACAGGATTAGTT	780
Qy	318	GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu	337
Db	781	CAAAATTTTCCATAACAATAATCAAAATTAATTGAAGGAAAAATCCAGATATGTTA	840
Qy	338	AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg	357
Db	841	AACATTTTCACAAACCAATCGCTPAAAAACAATGTCACAGAAAATTCGTGGATGTTTCAGA	900
Qy	358	HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys	377
Db	901	CATTTCAGTCAAAAGAGAAGAAATGTAATGTTTATAAATTACAAACAAGAGGTGATAAA	960
Qy	378	CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla	397
Db	961	TGTGTTGAAAATCCAAATCCCTACTTCACGAAAAATAATGTTGGATGTGTGATGCAGATGCC	1020
Qy	398	ThrCystThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys	417
Db	1021	ACATGTACCGAAGAAGATTCAGTGACCGACAGAAAGAAATCCACATGTGAATGTACTATAA	1080
Qy	418	ProAspSerTyrProLeuPheAspGlyIlePheCysSer	430
Db	1081	CCGTGATCTTATCCACTTTTCGATGTGATTTTCTGCAGT	1119

[illegible]


```

NNSVYVQK LKKALSYLEDYSLRGISEKDPNHYTTLKTGLEADIKKLTBEIKSSENK
ILEKRGFLTHSANALESVDYIVLQVQVLLIKKIEDLRKIELFLNAQKDKSIHVP
NIYKPNKPEPYIIVLKVKEVDKLEFIPKVKMLKKEQAVISSITOPLVAASETTID
GGHSTHTLSQSGETEETETETETETETETETETETETETETETETETETETET
ALTVTYLLKDEFLTKSYICHKYLIVSNSSMDOKLLEVNLTPPEENELKSCDPLDL
LFNTONIPANYSLYDSNNNDLOHLFPELYQKEMLYLHLKLEENHIAKLLERQKQIT
GTSSTSPGNTVTWQASATHNSQNSQNSASSTNTQNGVAVSSGPVAVESHDPITV
LSISNDKIVGLNLGNKTKVPNPLTISTTEMEKFYENILKNNDITFNDDIKQFVK
NSKVITGLTETQKVALNDEIKKDLTQLSFDLYNKYKLLDLRLFNKKKELGQDKMQI
KKLLKLEQESKLSNNPNHVNQNFVFNKKKEAEIAETENTLENTLILKHKYKG
LVKYNGESSPLKTSVSIOTEDYANLEKFRVLSKIDGKLDNHLGKKKLSFLSS
GLHOLITJELKEVINKNATYGNPSNNKKVNEALKSYENELPEAKVTYVTPPOPDVT
PSPLSRVSGSGSTKETQIPTSGSLLELQQVVLQNDDEEDSLVLPIDFVSED
NDEYLDQVVTGEATSVTMDNLSGFENEYDVIYIKLAGVYRSLKQKOEKNIFFNLN
LDNLNLRKFLDYLDVLEDSLMQFSDHYSNEYIIEDSFKLNSEQKNTLLKSYKI
KESYENDIKFAEQESYKYLAKYKDDLESIKKVIKEKEKPSPPPTPPSPAKTD
EOKKESFLPPLTNIETLYNNLVNKIDYILNLKAKINDCNVEKDEAHVKTLLSDLK
AIDDKIDLKFNHNDPDAIKLINDTKMDLGLKLLSLGLVONFPNTILSKIEKQFD
MLNISOQCVMKKQCPENSGCFRHDERBECKLNLKQEGDKCVENPNPCNENNGGC
DADAKTEEDSGSGNGKITCCTKPDSPYLPFDGIFCSSNPLGISFLLILMLILYSFI
BASE COUNT      2325 a      686 c      711 g      1554 t
ORIGIN

Alignment Scores:
Pred. No.:      3,11e-126      Length:      5276
Score:          1932.50      Matches:      371
Percent Similarity: 96.45%      Conservative: 9
Best Local Similarity: 94.16%      Mismatches: 9
Query Match:      84.95%      Indels:      5
DB:              3          Gaps:      1

US-10-057-531a-2 (1-431) x PFAGP195 (1-5276)

Qy 38 ThrAspAspAspValAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3952 TCCGAAGATAATGACGATAATTAGCATCAAGTAGTAACTGGAGNA----- 3996

Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3997 GCAATATCTGTCACAAATGGAATAATCTCTCAGGATTGGAATAATGAATGATGTTATA 4056

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4057 TATTTAAACCTTTACGTGGAGTATATAGACCTTAAAAAACAATTTGAAAAACAATT 4116

Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4117 TTTACATTTAAATTTAAATTTGACCATATCTTAATTCACGCTTTAAGAACGAAATAT 4176

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4177 TTCTTAGATGATTAGATCTGATTAAATGCAATTTAAACATATATATCTCAATCAATAC 4236

Qy 138 IleIleGluAspSerPheLysLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4237 ATTATTGAAGATTCAATTTAAATTTAGATTGCAACAAAAAACAACACTTTTAAAAAGT 4296

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4297 TACAATATATAAAGATCAGTACAAATGATATAATTTGACACGAGGTATAGT 4356

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysValIle 197
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4357 TATTATGAAAAGGTTTATAGCAAAATATAGGATGATTTAGAAATCAATTTAAAAAGTTATC 4416

Qy 198 LysGluGluLysGluLysPheProSerProThrThrProThrProSerProAlaLys 217
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4417 AAAGAAGAAAAGGAGAGTCCCATCATCACCAACCAACACCTCCCGTCACCAAGCAAAA 4476

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4477 ACAGACGACAAAGAAAGAAAGTAAGTTCCCTCCATTTTAAACAAACATTTGAGACCTTA 4536

```

```

Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrIleAsnLeuLysAlaLysIleAsn 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4537 TACAATACTTAGTTAAATAAATTCACGATTACTTAATTAATTAACGCAAGATTAAC 4596

Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4597 GATTGTAATGTTGAAAAAGATGAACACATGTTTAAATAAATTAACACTTAGTGATTTAAA 4656

Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaLysLys 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4657 GCAATGTGACAAAATAGATCTTTTAAAAACCAATACGACTTCGATGCAATTTAAAAA 4716

Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4717 TTGATAAATGATGATACGAAAAAAGATATGCTTGGCAAAATTTACTTAGTACAGGATTAGT 4776

Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4777 CAAATTTTCTTAATACAAATATATCAAAATTAATTTGAAGGAAAAATTCACAGATATGTTA 4836

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4837 AACATTTTCAACACCAATGCGTAAAAAACAATGTCAGAAAAATTCGGATGTTTCAGA 4896

Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluLysAspLys 377
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4897 CATTTAGATGAAAGAGAGAATGTAATGTTTATTAATTAATTAACAAAGAGTGATATAA 4956

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4957 TGTGTTGAAAATCCAAATCTACTTGTAAACGAAAAATTAATGTTGGATGTCAGTACATGCC 5016

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5017 AAATGTCGGAAGAGATTTCAGGTAGCAACGAGAAAGAAATACACATGTGAATGTACTAAA 5076

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5077 CCTGATCTTATCCACTTTTTCGATGTTATTTCTGCAGTTCC 5118

RESULT 15
PFAMSP8221
LOCUS      1726 bp      DNA      linear      INV 29-MAY-2002
DEFINITION P. falciparum DNA for the precursor to the major merozoite surface
            proteins, C-terminal.
ACCESSION  D13347
VERSION    1 GI:391794
KEYWORDS   EGF-like domains; major merozoite surface protein precursor.
SOURCE     Plasmodium falciparum (isolate 822/1) merozoite, DNA.
ORGANISM   Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE  1 (bases 1 to 1726)
            Jongwutives S., Tanabe K. and Kanbara H.
            Sequence conservation in the C-terminal part of the precursor to
            the major merozoite surface proteins (MSP1) of Plasmodium
            falciparum from field isolates
            Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
            93295445
            2 (bases 1 to 1726)
            Jongwutives S.
            Direct Submission
            Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical
            Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4
            Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747),
            Fax:0958-47-6607)
            Submitted (05-OCT-1992) to DBJ by:
            Somchai Jongwutives
            Department of Protozoology
            Institute of Tropical Medicine
            Nagasaki University
            12-4 Sakamoto-machi
            Nagasaki 852

```

```

US-10-057-531A-2 (1-431) x PFAMSPB221 (1-1726)
Qy 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57
Db 486 TCCGAAGATGATGACGAATATTATTAGATCAAGTAGTACGTGCAGAA----- 530
Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
Db 531 GCAATATCTGTCACAAATGGATAATATCCTCTCAGGATTTGAAAATGAATATGATGTTATA 590
Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
Db 591 TATTATAAACCTTTAGCTGGNGATATAGAAGCTTAAAAAACAATTTGAAAAACACTT 650
Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 651 ATTTACATTTAATTTAAATTTGAACGATATCTTAAATTCACGCTCTTAAGAAACGAAAAATAT 710
Qy 118 PheLeuAspValLeuGluSerAspLeuMetClnPheLysHisIleSerSerAsnGluTyr 137
Db 711 TTCCTTAGATGATTAGAAATCTGATTTAATGCAATTTAAACATATATCCTCAAAATGAATAC 770
Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
Db 771 ATATTAGACATTCATTTAAATTTAGATTTACAGACAAAAAACAACACTTTTAAAAAGT 830
Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
Db 831 TACAAATATATAAAAGGAATCAGTAGAAAAATGATATTAATTTGCACAGCAAGGTATAAGT 890
Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle 197
Db 891 TATTATGAAAGGTTTATAGCGAAATATAAAGGATGATTAGAAATCAATTTAAAAAAGTTATC 950

```



```

REFERENCE 2 (bases 1 to 1726)
AUTHORS  Jongwutwises,S.
TITLE      Direct Submission
JOURNAL    Submitted (05-OCT-1992) Somchai Jongwutwises, Institute of Tropical
           Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4
           Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747)),
           Fax:0958-47-6607)
COMMENT    Submitted (05-OCT-1992) to DDBJ by:
           Somchai Jongwutwises
           Department of Protozoology
           Institute of Tropical Medicine
           Nagasaki University
           12-4 Sakamoto-machi
           Nagasaki 852
           Japan
           Phone: 0958-47-2111x3747
           Fax: 0958-47-6607.
FEATURES   Location/Qualifiers
            source
              1..1726
              /organism="Plasmodium falciparum"
              /db_xref="taxon:5833"
              /chromosome="9"
              /dev_stage="merozoite"
            gene
              1..1712
              /gene="MSP1"
            CDS
              <1..1712
              /gene="MSP1"
              /codon_start=3
              /product="major merozoite surface protein precursor"
              /protein_id="BAA02615.1"
              /db_xref="GI:391819"
              /translation="PLKLTSEVSIQEDNYANLEKFRVLSKIDGKLNLDHLGKKKLS
              FLSSGLHLITELKEVINKNKTGNSPENNKVNEALKSYENFLPEAKVTVVTPQ
              PDVTPSPVSGSGSTKEETQIPTSGSLTLEQQVQLONYDEEDSDLVLPFG
              ESDNDEYLDVVTGEATSDMTNLSGFENEDVYILKPLAGVYRSKKOIEKNLIT
              ENLNDLISRLKRRKVFYLDLSDLMQFKHISNEYIIEDSFKLLSSEKRNLLKS
              KYIKESVENDIKAQEGISYERVLAKIKODLESIKKAVIEKEFFSPSPPTPPSPA.
              KTDQKESKLEPFLNLTETLNNLKNIDDDYLNKAKINDCNVKEDEAHVITKLS
              DLKADDKIDKNTNDPEAIKLNDDTKKMLGKLLSTGLVQNFPTNIIISKIEGK
              FQDMLNTSHQCVRKQCPENSGCFRHLDERECCLLNKKOEGKQCVNPNPTNENN
              GCGDADATCTEEDSGSSRRKKTCECTKPDSPVLPFDGIFCSSNFIIGISFLILMLLY
              SFI"
BASE COUNT 740 a 233 c 245 g 508 t
ORIGIN
Alignment Scores:
Pred. No.: 1,23e-126 Length: 1726
Score: 1931.00 Matches: 373
Percent Similarity: 96.45% Conservative: 7
Best Local Similarity: 94.67% Mismatches: 8
Query Match: 84.88% Indels: 6
DB: 3 Gaps: 2
US-10-057-531A-2 (1-431) x PFAMSP946 (1-1726)
QY 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57
      ::::::::::::::: :::: :::: ::::
DB 486 TCCGAAGATAATCAGCAATATTATGATCAAGTAGTACGAGAA----- 530
QY 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
DB 531 GCATATCTGTCACAAATGGAATAATCCCTCTCAGGATTTGAAATGAATATGATGTATA 590
QY 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
DB 591 TATTTAAAACCTTAGCTCGATATATAGACCTTAAAAAACAATTCGAAAAACATT 650
QY 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
DB 651 ATTACATTTAATTAATTTGAACGATATCTTAAATTCAGCTCTTAAGAAACGAAATAT 710
QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::

```

```

Db 711 TTCTTAGATGTATTAGAATCTGATTAAATGCAATATATAATATATATCTCAATGAATAC 770
QY 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
Db 771 ATATTGAAGATTCATTAAATATTGAATTCAGAACAAAAACACACTTTTAAANAAGT 830
QY 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
Db 831 TACAAATATATAAAGAATCAGTAGAAAATGATATTAATTTTGCACAGGAAGTATAAGT 890
QY 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysValIle 197
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
Db 891 TATTATGAAGAGTTTACCGAAATATAAGGATGATTTAGAATCAATTTAAAAAAGTATC 950
QY 198 LysGluGluLysGluLysPheProSerSerProThrThrProProSerProAlaLys 217
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
Db 951 AAAGAAGAAAAGAGAG---TTCCCATCATCCACCACCAACACCTCCGTCACGACGAAA 1007
QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
Db 1008 ACAGACGAACAAAAGAGAAAGTAAGTTCTTCCTTCCATTTTAAACAACATTCGAGACCTTA 1067
QY 238 TyrAsnAsnLeuValAsnLysLysLeuAspAspTyrLeuIleAsnLeuLysAlaLys 257
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
Db 1068 TACAATACTTAGTTAATAAATAATGACGATTACTTAAATTAACCTAAAGCAAGATTAAAC 1127
QY 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
Db 1128 GATTGTATGTTGAAAAGATGACGACATGTTAAATAACTAAACTTAGTATTTAAAA 1187
QY 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaLysLys 297
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
Db 1188 GCAATTGTATGACAAAATAGATCTTTTAAAAACACTAACGACTTCGAGCAATTTAAAAA 1247
QY 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
Db 1248 TTGATAAATGATGATACGAAAAGAGATATGCTTGGCAAAATACTTAGTACAGGATTAGTT 1307
QY 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
Db 1308 CAAATTTTCTTAATACATAATATACAAATTAATTTGAAGGAAAAATTCACAGATATGTTA 1367
QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
Db 1368 AACATTTTCAACACCAATCGCTAAAAAACAATGTCCAGAAAAATTTCTGGATGTTTCAGA 1427
QY 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
Db 1428 CATTTAGATGAAGAGAGAAATGTAATGTTTATTAAATTACAAACAAGAGGTGATAAA 1487
QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlnGlyCysAspAlaAspAla 397
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
Db 1488 TGTGTTGAAAATCCAAATCCTACTTGTAAACGAAAAATAATGGTGGATGTCAGATGCC 1547
QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
Db 1548 ACATGTACCGAGAGAGATTCAGGTAGCAGACAGAAAATAACATCATGTGAATGTACTAAA 1607
QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
      ::::::::::::::: ::::::::::::::: ::::::::::::::: :::::::::::::::
Db 1608 CCTGATCTTATCCACATTTTCGATGTTATTTTCTGCAGTTCC 1649

```

```

RESULT 18
LOCUS      AF040568
DEFINITION Plasmodium falciparum strain Kenya-2 merozoite surface protein-1
            (MSP-1) gene, partial cds.
ACCESSION AF040568
VERSION    AF040568.1
KEYWORDS   GI:2789667
SOURCE     Plasmodium falciparum.
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

```

REFERENCE 1 (bases 1 to 1119)
 AUTHORS Qari,S.H., Shi,Y.P., Goldman,I.F., Nahlen,B.L., Tibayrenc,M. and Lal,A.A.
 TITLE Predicted and observed alleles of plasmodium falciparum merozoite surface protein-1 (MSP-1), a potential malaria vaccine antigen
 JOURNAL Mol. Biochem. Parasitol. 92 (2), 241-252 (1998)
 MEDLINE 98319411
 PUBMED 9657329
 REFERENCE 2 (bases 1 to 1119)
 AUTHORS Qari,S.H., Shi,Y.P., Goldman,I., Nahlen,B., Tibayrenc,M. and Lal,A.A.
 TITLE Direct Submission
 JOURNAL Submitted (31-DEC-1997) DASTUR, Centers for Disease Control & Prevention (CDC), 1600 Clifton Road, Atlanta, GA 30333, USA
 FEATURES
 source
 1. .1119
 /organism="Plasmodium falciparum"
 /strain="Kenya-2"
 /isolate="20"
 /db_xref="taxon:5833"
 /chromosome="9"
 <1. .>1119
 /gene="MSP-1"
 /allele="Kenya-2"
 <1. .>1119
 /gene="MSP-1"
 /note="42 kDa C-terminal region"
 /codon_start=1
 /product="merozoite surface protein-1"
 /protein_id="AAC39098.1"
 /db_xref="GI:2789668"
 /translation="AIVSTMDNLSGFENEYDVIYKPLAGVYRSKKQIEKNITFN
 LNLINDLSRKKRYFLDLESLDMQFKHISSEYIIEDSFKLNSEQKNTLLKSYK
 YIKESVENDIKFAQEGISYEVKLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAK
 TDQKESKFLPPLTNIETLYNNLVNKIDYLLNLKAKINDCNVDEAHVKTIKLSD
 LKAIIDKIDLFKNHNPBEAKKLINDTKDKMLGKLLSLGLVQNFNTIISKLEKFP
 QDMNLSQRCVKRQCFENSGRCRHDRECKLLNYKQEGDKCVENPNPTCNENNG
 GCDADAKTEEDSGSRKKITCECTKSYSYPLFDGFC"
 BASE COUNT 487 a 144 c 162 g 326 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.08e-126 Length: 1119
 Score: 1929.00 Matches: 369
 Percent Similarity: 98.93% Conservative: 0
 Best Local Similarity: 98.93% Mismatches: 4
 Query Match: 84.79% Indels: 0
 DB: 3 Gaps: 0

US-10-057-531A-2 (1-431) x AF040568 (1-1119)

Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
 |||||
 Db 1 GCAATATCTGTCAACATGGATAATATCTCTCAGGATTTGAAATGAATGATGTTATA 60
 |||||

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
 |||||
 Db 61 TATTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAAACAATAATGAAAAACATT 120
 |||||

Qy 98 PheThrPheAsnLeuAsnLeuAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
 |||||
 Db 121 TTTACATTTAATTTAAATTTGAACGATATCTTAATTCAGCTCTTAAGAACGAAATAT 180
 |||||

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
 |||||
 Db 181 TTCTTAGATGATTAGATCTGATTTAATGCAATTTAAACATATATCTCAATGATAC 240
 |||||

Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSer 157
 |||||
 Db 241 ATTATTGAAGATTCATTTAAATTTAATTCAGAACAAAAAACACACTTTTAAAAAGT 300
 |||||

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
 |||||

Db 301 TACAAATATATAAAGAAGTACAGTAGAAAAATCATATTAAATTTGCACAGGAAGGTATAGT 360
 Qy TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle 197
 |||||

Db 361 TATTATCAAAAGGTTTGTAGCGAAATATATAGGATGATTTAGATCAATATAAAGAGTTATC 420
 |||||

Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
 |||||

Db 421 AAACAACAAAAAGGAGAAGTTCCCATCATCATCACCAACAACACCTCGCTCACCAGCAAAA 480
 |||||

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
 |||||

Db 481 ACAGACCAACAAAAAGAGAAAGTAAGTTCTTCCATTTTAAACAACATTGAGACCTTA 540
 |||||

Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
 |||||

Db 541 TACAATAACTTAGTTAATAAAATTGACGATTACTTAATTAACCTTAAGGCAAGATTAAAC 600
 |||||

Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
 |||||

Db 601 GATTGTAATGTTGAAAAAGATGAAGCACATGTTAAAAATACTAAACTTAGTGATTAAAA 660
 |||||

Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
 |||||

Db 661 GCAATTGATGACAAATAGATCTTTTAAAAACCATTAACGACTTTCGAAGCAATTTAAAAA 720
 |||||

Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
 |||||

Db 721 TTGATAAATGATGATGACAAAAAGATATGCTTGGCAAAATTACTTAGTACAGGATTAGTT 780
 |||||

Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluLysPheGlnAspMetLeu 337
 |||||

Db 781 CAAAATTTCTTAATCAACAATAATCAAAATTAATTAAGGAAAAATTTCCAAGATATGTTA 840
 |||||

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGluCysProGluAsnSerGlyCysPheArg 357
 |||||

Db 841 ACAATTTCAACAACCACTCGTAAAAAACAATGTCCAGAAAAATCTCGGATGTTTTCAGA 900
 |||||

Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
 |||||

Db 901 CATTTAGATCAAGAGAGAAGTGAATGTTAATTTAATAATTAACAACAAGAGGTGATAAA 960
 |||||

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
 |||||

Db 961 TGTGTTGAAAAATCCAAATCTCTACTTGTAAAGAAAAATTAATGTTGATGATGCGAGATGCC 1020
 |||||

Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
 |||||

Db 1021 AATGTACCCAGAGAGATTCAGGTAGCAGCAGCAAGAAAAATCATCATGTGAATGTACTAAA 1080
 |||||

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
 |||||

Db 1081 TCTGATTTCTTATCCACTTTTCGATGGTATTATTCTGCAGT 1119
 |||||

RESULT 19
 PFU20733
 LOCUS
 DEFINITION PFU20733 1119 bp DNA linear INV 14-FEB-1995
 Plasmodium falciparum clone B17D merozoite surface protein 1
 (MSP-1) gene, 42 kDa C-terminal region, partial cds.
 ACCESSION U20733
 VERSION U20733.1 GI:664867
 KEYWORDS
 SOURCE malarial parasite.
 ORGANISM Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 1119)
 AUTHORS Shi,Y.-P., Alpers,M.P., Povoa,M.M., Nahlen,B.L., Oloo,A.G. and Lal,A.A.
 TITLE Sequence of the C-terminal region of merozoite surface protein-1 (MSP-1) in field-derived Plasmodium falciparum
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1119)
 AUTHORS Shi,Y.-P.

TITLE Direct Submission
JOURNAL Submitted (07-FEB-1995) Ya-Ping Shi, DPD/CDC 4770, Molecular Vaccine Section, Bldg 22, Rm. 8F12, Buford Highway, Chamblee, GA, 30341 USA

FEATURES
source Location/Qualifiers
1. .1119
/organism="Plasmodium falciparum"
/isolate="infected human blood sample from Papua New Guinea"
/db_xref="taxon:5833"
/clone="B17D"
1. .1119
/gene="MSP-1"
/c1.>.1119
/gene="MSP-1"
/codon_start=1
/product="merozoite surface protein 1, 42 kDa C-terminal region"
/protein_id="AAA62224.1"
/db_xref="GI:664868"
/translation="AISVTMDNLSGFENEYDVIYKPLAGVRSILKKOIEKNITFN LNLNLSRLKRRKRVFELDVLEDFMOFKHISSEYIIEEDSKLLNSEKNTLLKSYK YKESVENDIKFAQEGISYEVKVLAKYKDDLESIKKVIKEEKEPPSPPTPPSPAK TDEQKESKFLPFLNIEFLYNNLVKIDDIYLNKAKINDCNVKNENHAKITKLSD LKAIIDKIDLFKNTDFEAIKKLINDTKKMDLGLLSTGLVONFPNTIISKLEGEK ODLNISOHCYKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPDPTCENNG GCDADATCTEEDSGSSRRKITCECTKPDSPYPLFDGIFCS"

BASE COUNT 487 a 146 c 162 g 324 t

ORIGIN

Alignment Scores:
Pred. No.: 1.27e-126 Length: 1119
Score: 1928.00 Matches: 368
Percent Similarity: 99.20% Conservative: 2
Best Local Similarity: 98.66% Mismatches: 3
Query Match: 84.75% Indels: 0
DB: 3 Gaps: 0

US-10-057-531A-2 (1-431) x PFU20733 (1-1119)

QY 58 AlaileSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
|||||
DB 1 GCAATATCTGTCACATGGATAATATCTCTCAGGATTTGAAATGAATATGATGTTATA 60
QY 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
|||||
DB 61 TATTTAAACCTTTAGCTGAGTATATAGAAAGCTTAAACAAACAAATTCGAAAAACATT 120
QY 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
|||||
DB 121 ATTACATTTAAATTTAAATTTGAACGATATCTTAAATTCAGCTTAAAGAAACGAAATAT 180
QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
|||||
DB 181 TTCTTAGATGATTAGAAATCGATTAAATGCAATTTAAACATATATCCCAAAATGAATAC 240
QY 138 IleileGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
|||||
DB 241 ATTATGAAGATTCATTTAAATTTATTTGAATTCAGACACAAAAACACACTTTTAAAGT 300
QY 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
|||||
DB 301 TACAAATATATAAAGAATCAGTAGAAATGATATTTAAATTTTCACAGGAAGGTATAAGT 360
QY 178 TyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValIle 197
|||||
DB 361 TATTATGAAGAGGTTTACGAAATATAAGGATGATTGAATCAATTTAAAGGATTCATC 420
QY 198 LysGluGluLysLysPheProSerSerProProThrThrProProSerProAlaLys 217
|||||
DB 421 AAAGAGAAAAGGAGAGGTTCCCATCATCACCACCAACAAACACCTCCGTCACGACGAAA 480
QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
|||||

Db 481 ACAGACGAACAAAAGAGAAAGTAAAGTTCCTCCATTTTAAACAAACATTGAGACCTTA 540
QY 238 TyrAsnAsnLeuValasnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
Db 541 TACAATAACTTAGTTAATAAATTGACGATTACTTAATAACTTAAAGGCAAGATTAAC 600
QY 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
Db 601 GATTGTANTGTTGAAAAAATGAGCACATGTTAAATAACTAAACTTAGTGTATTTAAAA 660
QY 278 AlaileAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
Db 661 GCAATTGATGACAAAAATAGATCTTTTAAACACACTAACGACTTCGAAGCAATTTAAAA 720
QY 298 LeuileAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
Db 721 TTGATAAATGATGATACGAAAAAAGATATGCTTTGGCAAAATTACTTAGTACAGGATTAGT 780
QY 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluLysPheGlnAspMetLeu 337
Db 781 CAAAATTTCTTAATAATAATATCAAAATTAATTGAAGAAAAATTCGAAGATATGTTA 840
QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 841 AACATTTCAACACCAATCGCTAAAAAACAATGTCCAGAAAAATTTCTGGATGTTTCAGA 900
QY 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
Db 901 CATTTAGATGAAAGAGAAGAAATGTAATGTTTAAATTTACAAACAAGAGGTGATAAA 960
QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397
Db 961 TGTGTTGAAATTCAGATCTCTACTTGTACGAAAAAATTAATGTTGATGTCAGATGCC 1020
QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
Db 1021 ACATGTACCGAAGAAGATTTCAGGTAGCAGCAAGAAAAATCATGTGTAATGACTATAA 1080
QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
Db 1081 CCTGATCTTTATCCACTTTTCGATGGTATTTCTTCGAGT 1119

RESULT 20
PFU20729 1119 bp DNA linear INV 14-FEB-1995
LOCUS Plasmodium falciparum clone K47D merozoite surface protein 1
DEFINITION (MSP-1) gene, 42 kDa C-terminal region, partial cds.
ACCESSION U20729
VERSION U20729.1 GI:664859
KEYWORDS malarial parasite.
SOURCE Plasmodium falciparum
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
REFERENCE 1 (bases 1 to 1119)
AUTHORS Shi,Y.-P., Alpers,M.P., Povoa,M.M., Nahlen,B.L., Oloo,A.G. and Lal,A.A.
TITLE Sequence of the C-terminal region of merozoite surface protein-1 (MSP-1) in field-derived Plasmodium falciparum
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1119)
AUTHORS Shi,Y.-P.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-1995) Ya-Ping Shi, DPD/CDC 4770, Molecular Vaccine Section, Bldg 22, Rm. 8F12, Buford Highway, Chamblee, GA, 30341 USA
FEATURES
1. .1119
Location/Qualifiers
/organism="Plasmodium falciparum"
/isolate="infected human blood sample from Papua New Guinea"
/db_xref="taxon:5833"
/clone="K47D"


```
gene      1. .1119
CDS       <1. .>1119
          /gene="MSP-1"
          /gene="MSP-1"
          /codon_start=1
          /product="merozoite surface protein 1, 42 kDa C-terminal
          region"
          /protein_id="AAAG2220.1"
          /db_xref="GI:664860"
          /translation="AISVTMDNLTSGFENEYDVYLKPLAGVYRSLKQIEKNLITFN
          LNLNDILNSLRKRYKFLVDLESDLMQFKHISNEYIIESFKLLNSEQNTLLKSVK
          YIKESVENDIKFQEGISYEKVLAKYKDDLESIKRKVEKEKFPSPPTPPSPAK
          TDEQKESKLPETLLETLYNNLNKIDYLLNLKAKINDCNVDEAHVITKLSLD
          LKAIDDDLFKHNDEFAIKLLINDTKKMLGKLLSTGLVQNFPTIISKIEGRF
          QDMLNIHQVCVKKQPCQNSGCFRHLDEREECKLLNKQEGDKCVENPNPTCNENNG
          GCDADAKCTEEDSGSKKITCECTKPSDYPLFDGIFCS"
BASE COUNT 487 a 146 c 161 g 325 t
ORIGIN

Alignment Scores:
Pred. No.:      2,41e-126      Length:      . 1119
Score:          1924.00      Matches:      . 367
Percent Similarity: 98.93%      Conservative: 2
Best Local Similarity: 98.39%      Mismatches: 4
Query Match:      84.57%      Indels:      0
DB:              3          Gaps:      0

us-10-057-531a-2 (1-431) x PFU20729 (1-1119)

Qy 58 AlalieserValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValile 77
    |||||
Db 1 GCAATATCTGTCAATGGATAATATCTCTCAGGATTTGAAATGAATATGATGTTATA 60

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
    |||||
Db 61 TATTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAAACCAATTTGAAAAACATT 120

Qy 98 PheThrPheAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
    |||||
Db 121 TTTACATTTAATTTTAAATTTGAACGATATCTTAAATTCACGCTCTTAAGAAACGAAATAT 180

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
    |||||
Db 181 TTCTTAGATGTATTAGATCTGATTTAATGCAATTTAAACATATATCTCAATGAATAC 240

Qy 138 IleLeuGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
    |||||
Db 241 ATTATTGAAGATTTCATTTAAATTTTGAATTCAGAACAAAAAACACACTTTTAAAAAGT 300

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
    |||||
Db 301 TACAAATATATAAAGAAATCAGTAGAAATGATATTAATTTGCACAGGAAGGTATAAGT 360

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysValile 197
    |||||
Db 361 TATTATGAAAGGTTTAGCCAAATATAAGATGATTTAGATCAATCAATTAATAAGTTATC 420

Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
    |||||
Db 421 AAAGAAGAAAAAGGAGAGTTCCCATCATCACCACCAACACACCTCCGTCACCAAGCAAAA 480

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
    |||||
Db 481 ACACAGCAACAAAAAGAGAAAGTAAGTTCCCTTCATTTTAAACAAACATTGAGACCTTA 540

Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
    |||||
Db 541 TACAATACTTACTTAATAAAATTTGACGATTACTTAATTAATTAAGGCAAAATTAAC 600

Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
    |||||
Db 601 GATTGTAATGTTGAAAAAGATGAAGACATGTTAAAAATACTAACTAACTTAGTGATTAAAA 660
```

```
Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
    |||||
Db 661 GCAATTGATGACAAAATAGATCTTTTTAAAAACCATACGACTTCGAAGCAATTAATAAAA 720

Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
    |||||
Db 721 TTGATAAATGATGATACGAAAGAAAGATATGCTTGCAAAATTTACTTAGTACAGGATTAGTT 780

Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
    |||||
Db 781 CAAAATTTCTTAATACATATATCAAAATTAATTAAGGAAATTTCCAAGATATGTTA 840

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
    |||||
Db 841 AACATTTCAACAACCAATGCGTAAAAAAACAATGTCCACAAAATCTCGATGTTTCAGA 900

Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
    |||||
Db 901 CATTTAGATGAAGAGAGAAATGTAATGTTTATTAATTAACAACAAGAGGTGATAAA 960

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397
    |||||
Db 961 TGTGTTGAAATCCAAATCCTACTTGTAAACGAAATAATGTTGGATGTCAGATGCC 1020

Qy 398 ThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys 417
    |||||
Db 1021 AAATGTCACCGAAGAGATTTCAGTAGCAACGGAAGAAAATCACATGTGAATGTACTAAA 1080

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
    |||||
Db 1081 CCTGATCTTATTCACCTTTTCGATGGTATTTTCGCACT 1119

RESULT 21
PFU20653
LOCUS      PFU20653      1116 bp      DNA      linear      INV 14-FEB-1995
DEFINITION Plasmodium falciparum clone P24B merozoite surface protein 1
            (MSP-1) gene, 42 kDa C-terminal region, partial cds.
ACCESSION U20653
VERSION    020653.1
KEYWORDS   GI:664843
SOURCE     malarial parasite.
ORGANISM   Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE  1 (bases 1 to 1116)
            Shi, Y.-P., Alpers, M.P., Povoa, M.M., Nahlen, B.L., Oloo, A.G. and
            Lal, A.A.
            Sequence of the C-terminal region of merozoite surface protein-1
            (MSP-1) in field-derived Plasmodium falciparum
            Unpublished
            2 (bases 1 to 1116)
            Shi, Y.-P.
            Direct Submission
            Submitted (06-FEB-1995) Ya-Ping Shi, Molecular Vaccine Section,
            Bldg 22, Rm. 8F12, DPD/CDC 4770, Buford Highway, Chamblee, GA 30341
FEATURES
            source
            1. .1116
            /organism="Plasmodium falciparum"
            /isolate="infected human blood sample from Papua New
            Guinea"
            /db_xref="taxon:5833"
            /clone="P24B"
            1. .1116
            /gene="MSP-1"
            <1. .>1116
            /gene="MSP-1"
            /codon_start=1
            /product="merozoite surface protein 1, 42 kDa C-terminal
            region"
            /protein_id="AAAG2221.1"
            /db_xref="GI:664844"
            /translation="AISVTMDNLTSGFENEYDVYLKPLAGVYRSLKQIEKNLITFN
            LNLNDILNSLRKRYKFLVDLESDLMQFKHISNEYIIESFKLLNSEQNTLLKSVK
            YIKESVENDIKFQEGISYEKVLAKYKDDLESIKRKVEKEKFPSPPTPPSPAKT
```


DEQKESKELPFLTNITLYNNLVNKIDYILINKAKINDCNVEKDEAHVKITKILSDL
 RAIDDKIDFKNTNDFEAIKKLINDTTRKMLGKLLSTGLVONFPNTIISKLEIGKQ
 DMLNTSOHCYKOCPCNSGCFRHLDERECKLLNKOEGDKCVENPNPTCNENGG
 CDADATCTEEDSGSRKKITCECTKPDSPYPLFDGIFCS"

BASE COUNT 485 a 146 c 161 g 324 t
 ORIGIN

Alignment Scores:

Pred. No.: 3,07e-126 Length: 1116
 Score: 1922.50 Matches: 369
 Percent Similarity: 98.93% Conservative: 0
 Best Local Similarity: 98.93% Mismatches: 3
 Query Match: 84.51% Indels: 1
 DB: 3 Gaps: 1

US-10-057-531A-2 (1-431) x PFU20653 (1-1116)

QY 58 AlaIleSerValThrMetAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
 Db 1 GCAATATCTGTCAAAATGGATATATCTCTCAGGATTTGAAATGAATATGATGTATA 60
 QY 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
 Db 61 TATTTAAAACCTTTAGCTGAGTATATAGAGCTTAAAAAACAAATTTGAAAAACATT 120
 QY 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
 Db 121 ATTACATTTTAAATTTGAACGATATCTTAAATTCAGCTTAAAGAAACGAAATAT 180
 QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
 Db 181 TTCTTAGATGTATAGAAATCTGATTTAATGCAATTTTAAACATATATCTCAATGAATAC 240
 QY 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
 Db 241 ATTATTGAAGATTCATTTAAATATTGAATTCAGAACAAAAAACACACTTTTAAAGT 300
 QY 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
 Db 301 TACAATATATAAAAGAAATCAGTAGAAATGATATTAATTTGCACAGGAAGGTATAAGT 360
 QY 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle 197
 Db 361 TATTATTGAAGGTTTTACGAAATATAAGGATGATTAGATTCAATTTAAAAAAGTTATC 420
 QY 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
 Db 421 AAAGAAAGAAAGGAG---TTCCCATCATCTACCACCAACACACTCCGTCACCGCAAAA 477
 QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
 Db 478 ACAGACGAACAAAGAAAGAAAGTAAGTTCCCTTCCATTTTAAACAACATTGAGACCTTA 537
 QY 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
 Db 538 TACAATAACTTAGTTAATAAATGACGATTTACTTAATTAACCTTAAGCAAAAGATTAAAC 597
 QY 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
 Db 598 GATTGTANTGTTGAAAAAGATGAGCACATGTTAAAAATACTAATCTTAGTGTATTAATA 657
 QY 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaLysLysLys 297
 Db 658 GCAATTCATGACAAAATAGATCTTTTAAAAACACCTAACGACTTCGAAGCAATTAATAAA 717
 QY 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
 Db 718 TTGATAAATGATGATACGAAAAAGATATGCTTTGGCAAAATTACTTAGTACAGGATTAGT 777
 QY 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
 Db 778 CAAAATTTTCTTAATACAATAATATCAAAATTAATTTGAGGAAATTTCCAAGATATGTTA 837

QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
 Db 838 AACATTTCAACACCAATCGGTAAAAAACAATGTCCAGAAATTTCTGGATGTTTTCAGA 897
 QY 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
 Db 898 CATTTAGATGAAAGAGAAGAAATGTAATGTTTAAATTTACAAACAAGAGGTGATATA 957
 QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlnGlyCysAspAlaAspAla 397
 Db 958 TGTGTTGAAAAATCCAATCTTACTTGTACGAAAAATAATGTTGATGTGATGAGATGCC 1017
 QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
 Db 1018 ACATGTACCGAAGAAGATTTCAGTAGCAGCAGAAAGAAATCATGTGAATCTACTATA 1077
 QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
 Db 1078 CCTGATTCCTATCCACTTTTCGATGTTTCTGTCAGT 1116
 RESULT 22
 PFAMSP947
 LOCUS
 DEFINITION
 P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.
 PFAMSP947 1726 bp DNA linear INV 29-MAY-2002
 ACCESSION D13355
 VERSION GI:391820
 KEYWORDS EGF-like domains; major merozoite surface protein precursor.
 SOURCE Plasmodium falciparum (isolate 947) merozoite, DNA.
 ORGANISM Plasmodium falciparum
 REFERENCE 1 (bases 1 to 1726)
 AUTHORS Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 TITLE Jongwutiwes,S., Tanabe,K. and Kanbara,H.
 JOURNAL Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates
 MEDLINE Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
 REFERENCE 93295445
 2 (bases 1 to 1726)
 AUTHORS Jongwutiwes,S.
 TITLE Direct Submission
 JOURNAL Submitted (05-OCT-1992) Somchai Jongwutiwes, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)
 COMMENT Submitted (05-OCT-1992) to DDBJ by:
 Somchai Jongwutiwes
 Department of Protozoology
 Institute of Tropical Medicine
 Nagasaki University
 12-4 Sakamoto-machi
 Nagasaki 852
 Japan
 Phone: 0958-47-2111x3747
 Fax: 0958-47-6607.
 FEATURES
 source
 1. .1726
 /organism="Plasmodium falciparum"
 /db_xref="taxon:5833"
 /chromosome="9"
 /dev_stage="merozoite"
 1. .1712
 /gene="MSP1"
 <1. .1712
 /gene="MSP1"
 /codon_start=3
 /product="major merozoite surface protein precursor"
 /protein_id="BAA02616.1"
 /db_xref="GI:391821"
 /translation="PLKTLSEVSIOTEDNVANLEKFRALSKIDGKLNLDNLHLGKKKLS
 FLSSGLHLLITELKEVINKNKNTGNSPENNKKYNALASYNFLPEAKVTPVTPPQ
 PDVTPSLSVRVSGSGSKEETQIPTSGSLITELQGVVOLQNVDEEDSLVLPFG
 ESEDNDEYLDQVVTGTAISVTMDNILSGFENEYDVITYLKLAGVRSRLKKQIEKNIFT

LKAIDDKIDLFKNHNDFAIKKLIINDTKKMDLKLSTGLVQNPNTIISKLEIGKF
ODMLISQHYVKKOCPENSGCFRHLDERECKCLNLYKQEGDKCVENPNPTCENNG
GCDADAKCTEEDSGSGRKITCECTKPSYPLFDGIFCS"

BASE COUNT 487 a 145 c 162 g 325 t
ORIGIN

Alignment Scores:
Pred. No.: 5,41e-126 Length: 1119
Score: 1919.00 Matches: 368
Percent Similarity: 98.66% Conservative: 0
Best Local Similarity: 98.66% Mismatches: 5
Query Match: 84.35% Indels: 5
DB: 3 Gaps: 0

US-10-057-531A-2 (1-431) x AF040567 (1-1119)

QY 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
Db 1 GCAAATATCTGTCAACATGGATATATCTCTCAGGATTTGAAATGAATGATGTTATA 60
QY 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
Db 61 TATTTAAACCTTTAGCTGAGTATATAGAGCTTAAAAACAAATTCGAAANAACATT 120
QY 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 121 TTTACATTTTAAATTTGAACGATATCTTTAAATTCAGCTTAAAGAAACGAAATAT 180
QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
Db 181 TTTCTTAGATGATAGAAATCTGATTAATGCAATTTAAACATATATCTTCAATGAATAC 240
QY 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
Db 241 ATTATTGAAGATTCATTAAATATTGAATTCAGACACAAAAACACACTTTAAAGAGT 300
QY 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
Db 301 TACAAATATATAAAGAAATCAGTAGAAATGATATTAAATTTGCACAGGAAGGTATAAGT 360
QY 178 TyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValIle 197
Db 361 TATTATGAAGAGGTTTACGGAATATAGGATGATTAGATCAATTAAGAAAGTTATC 420
QY 198 LysGluGluLysLysPheProSerSerProProThrThrProProSerProAlaLys 217
Db 421 AAAGAAGAAAGGAGAGGTTCCCATCATCACCAACCAACACCTCCGTCACCAAGCAAAA 480
QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
Db 481 ACAGACGACAAAGAAAGGAAAGTAAAGTTCCCTTCCATTTTAAACAAACATTTGACACCTTA 540
QY 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
Db 541 TACAATAACTTAGTTAATAAATTGACGATTACTTAATTAACCTTAAGCGCAACATTAAAC 600
QY 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
Db 601 GATTGTGATGTTGAAAGATGAAGCACATGTTAAAAATACTAACTTAGTGATTTAAAA 660
QY 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
Db 661 GCAATTGATGACAAAAATAGATCTTTTAAAAACCATTAACGACTTCGAAGCAATTAAGAAA 720
QY 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
Db 721 TTGATAATATGATATACGAAAAAAGATATGCTTTGGCAAAATTACTTAGTACAGGATTAGTT 780
QY 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
Db 781 CAAAAATTTCTTAATACATAATATCAAAATTAATTAAGGAAAAATTCGAAGATATCTTA 840
QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357

Db 841 AACATTTACACACCAATACGTAAACAAACAAATGTCCAGAAATCTCGATGTTTCAGA 900
QY 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
Db 901 CATTTAGATGAAAGAGAAGAAATGTAATGTTTATTAATTAACAAACAAAGAGTGATAAA 960
QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397
Db 961 TGTGTTGAAATCCAAATCCTACTTGTACGAAAAAATAATGTTGATGATGATGATGCC 1020
QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
Db 1021 AAATGTACCGAAGAAGATTGAGTAGCAGCGGAAAGAAATACATGTTGAATGACTATAA 1080
QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
Db 1081 CTGTATTTTATCCATTTTCGATGATGATTTCTTCGAGT 1119
RESULT 24
PFU20726 1119 bp DNA linear INV 14-FEB-1995
LOCUS Plasmodium falciparum clone K46A merozoite surface protein 1
DEFINITION (MSP-1) gene, 42 kDa C-terminal region, partial cds.
ACCESSION U20726
VERSION U20726.1 GI:664853
KEYWORDS malarial parasite.
SOURCE Plasmodium falciparum
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 1119)
AUTHORS Shi,Y.-P., Alpers,M.P., Povoa,M.M., Nahlen,B.L., Oloo,A.G. and Lal,A.A.
TITLE Sequence of the C-terminal region of merozoite surface protein-1 (MSP-1) in field-derived Plasmodium falciparum
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1119)
AUTHORS Shi,Y.-P.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-1995) Va-Ping Shi, DPD/CDC 4770, Molecular Vaccine Section, Bldg 22, Rm. 8F12, Buford Highway, Chamblee, GA, 30341 USA
FEATURES
Source Location/Qualifiers
1. .1119
/organism="Plasmodium falciparum"
/isolate="infected human blood sample from Papua New Guinea"
/db_xref="taxon:5833"
/clone="K46A"
1. .1119
/gene="MSP-1"
<1. >1119
/gene="MSP-1"
/codon_start=1
/product="merozoite surface protein 1, 42 kDa C-terminal region"
/protein_id="AAA62217.1"
/db_xref="GI:664854"
/translation="AISVTMDNTLSGFENEYDVIYLLKPLAGVYRSLLKQIEKNIFTFN
LNNDLLNSRLKKRYFLDYLDLDMQFKHISSENVIIEDSPKLLNSEKNTLLKSYK
YIKESVDENIKFAQEGISYIEKYLAKYKODLESIRKVIKEEKFPSPPTPPSPAK
TDEQKESKFLPFLNIETLYNNLVNKIDHYLNKAKINDCNVKNDEARHVKILSD
LKAIDDKIDLFKNHNDFAIKKLIINDTKKMDLKLSTGLVQNPNTIISKLEIGKF
ODMLISQHYVKKOCPENSGCFRHLDERECKCLNLYKQEGDKCVENPNPTCENNG
GCDADAKCTEEDSGSGRKITCECTKPSYPLFDGIFCS"
BASE COUNT 486 a 146 c 162 g 325 t
ORIGIN
Alignment Scores:
Pred. No.: 5,41e-126 Length: 1119
Score: 1919.00 Matches: 366
Percent Similarity: 98.66% Conservative: 2
Best Local Similarity: 98.12% Mismatches: 5

```

Query Match: 84.35% Indels: 0
DB: 3 Gaps: 0
US-10-057-531A-2 (1-431) x PFU20726 (1-1119)

Qy 58 AlalleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValle 77
Db 1 GCAATATCTGTCACAAATGGATATATCTCTCAGGATTGAAAATGAATGATGTTATA 60

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIle 97
Db 1 TATTTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAAACAATTTGAAAAACAAT 120

Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 121 TTTACATTTAATTTAAATTTCAACGATATCTTAAATTCACGCTCTTAAGAAACGAAATAT 180

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
Db 181 TTCTTAGATGATTAGAAATCTGATTAAATGAATTTAAACATATATCTCAATGAATAC 240

Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSer 157
Db 241 ATTATTGAAGATTCATTTAAATTTAGATTTCAGAACAAAAACACACTTTTAAAAAGT 300

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
Db 301 TACAATATATAAAGAAATCAGTACAAATGATATTAATTTGCACAGGAAGGTATAAGT 360

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysLysValle 197
Db 361 TATTATGAAAGGTTTAAAGCAATATTAAGGATGATTTAGAAATCAATTTAAAAAGTTATC 420

Qy 198 LysGluGluLysGluLysPheProSerProProThrThrProProSerProAlaLys 217
Db 421 AAAGAAAGAAAGGAGAGTTCCTCATCATCACCAACACACCTCCGTCACGACAAAA 480

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
Db 481 ACACGACACAAAAAGAGAAAGTAAGTTCCTCATTTTAAACAAACATTTGAGACCTTA 540

Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
Db 541 TACAATACTTAGTTAATAAATTAAGCATTACTTAAATTAATTAAGGCAAGGATTAAC 600

Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
Db 601 CATTTGTAATGTTGAAAGAGATGAAGCAGATCTTAAATAACTAACTTAGCTGATTAAAA 660

Qy 278 AlalleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
Db 661 GCAATTGATGACAAAATAGATCTTTTAAAAAACCATACGACTTCGAAGCAATTAATAAAA 720

Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGluLysLeuLeuSerThrGlyLeuVal 317
Db 721 TTGATAATGATGATACAAAAAAGATATGTTGGCAATTTACTTAGTACAGGATTAGTT 780

Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluLysPheGlnAspMetLeu 337
Db 781 CAAATTTTCTATACATAATATCAAAATTAATTAAGGAAATTTCCAAGATATGTTA 840

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
Db 841 AACATTTTCACAAACCAATGCGTAAAAAACAATGTCCAGAAAAATTTCTGGATGTTTCAGA 900

Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluLysAspLys 377
Db 901 CATTTAGTAAAGAGAAAGATGTAATGTTTATTAATTAACAAACAGAGAGGTGATATAA 960

Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
Db 961 TGTGTTGAAATCCAAATCCTTACTTGTAAACGAAATTAATGTTGGATGTCAGATGCC 1020

Qy 398 ThrCysThrGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
Db 1021 AAATGTCACCGAAGAGATTAGTAGCAACGAAAGAAAGTACATGTAATGACTATAA 1080

Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
Db 1081 CCTGATCTTATCCACTTTTCGATGGTATTTTCGCAGT 1119

RESULT 25
PFU20654 1116 bp DNA linear INV 14-FEB-1995
DEFINITION Plasmodium falciparum clone P24C merozoite surface protein 1
(MSP-1) gene, 42 kDa C-terminal region, partial cds.
ACCESSION U20654
VERSION U20654.1 GI:664845
KEYWORDS malarial parasite.
SOURCE Plasmodium falciparum
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 1116)
AUTHORS Shi,Y.-P., Alpers,M.P., Povoa,M.M., Nahlen,B.L., Oloo,A.G. and Lal,A.A.
TITLE Sequence of the C-terminal region of merozoite surface protein-1
(MSP-1) in field-derived Plasmodium falciparum
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1116)
AUTHORS Shi,Y.-P.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-1995) Ya-Ping Shi, Molecular Vaccine Section,
Bldg 22, Rm. 8F12, DPD/CDC 4770, Buford Highway, Chamblee, GA 30341
FEATURES
source
/organism="Plasmodium falciparum"
/isolate="infected human blood sample from Papua New
Guinea"
/db_xref="taxon:5833"
/clone="P24C"
1..1116
/gene="MSP-1"
<1..>1116
/gene="MSP-1"
/codon_start=1
/product="merozoite surface protein 1, 42 kDa C-terminal
region"
/protein_id="AAAG62214.1"
/db_xref="GI:664846"
/translators="ATSVTMDNLLSGFENEVDVYLKPLAGVYRSLKQIEKNIIITEN
LNLDILNLRKKRKYFLDLESDLMQFKHISNEYIIEDESKLLNSEQNLLKSTK
YIKESVENDIKFAQEGISYVEKVLAKYDDLESIKKVIKEEFPPSPPTPPSPAKT
DEQKESKFLPFTNIETLYNNLVNKIDDDYILNLKAKINDCNVEKDEAHVYIKLSDL
KATDDKIDLEKNTNDFAIKKLINDDTKDMGLKLLSTGLVQNPNTIISKLEKFO
DMLNISOHCVKKOCPENSGCFRHLDERECKLLNLYKQEGDKCVENPSPTCNENNG
CDADATCTEEDSGSRKKITCECTKPSPYPLFDGIFCS"
BASE COUNT 484 a 145 c 162 g 325 t
ORIGIN
Alignment Scores:
Pred. No.: 6.87e-126 Length: 1116
Score: 1917.50 Matches: 368
Percent Similarity: 98.93% Conservative: 1
Best Local Similarity: 98.66% Mismatches: 3
Query Match: 84.29% Indels: 1
DB: 3 Gaps: 1
US-10-057-531A-2 (1-431) x PFU20654 (1-1116)
Qy 58 AlalleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValle 77
Db 1 GCAATATCTGTCACAAATGGATATATCTCTCAGGATTGAAAATGAATGATGTTATA 60
Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlnIleGluLysAsnIle 97
Db 1 TATTTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAAACAATTTGAAAAACAAT 120
TATTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAAACAATTTGAAAAACAAT 120

```

QY 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
 |||||
 Db 121 ATTACATTTAAATTTGAACGATATCTTAAATTCAGCTTAAAGAAACGAAATAT 180
 QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
 |||||
 Db 181 TTCTTAGATGTTATGAATCTGATTTAATGCAATTTAAACATATATCTCCCAATGAATAC 240
 QY 138 IleLeuGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSer 157
 |||||
 Db 241 ATTATTGAAGATTCATTTAAATTTGAATTCAGAACAAAAACACATTTTAAAGT 300
 QY 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
 |||||
 Db 301 TACAAATATATAAAGAAATCAGTAGAAAAATGATATTAATTTGCACAGCAAGGTATAAGT 360
 QY 178 TyrTyrGluValLeuAlaLysTyrLysAspAspLeuGluSerIleLysValIle 197
 |||||
 Db 361 TATTATGAAGAGTTTACGGAATATAAGGATGATTAGAAATCAATTTAAAAAGTTATC 420
 QY 198 LysGluGluLysGluLysPheProSerSerProThrThrProProSerProAlaLys 217
 |||||
 Db 421 AAAGAGAAAGAG--TTCCCATCATCACCACCAACACACCTCCGTCACCACGCAAA 477
 QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
 |||||
 Db 478 ACAGACGAAACAAAGAAAGTAAGTTCCTTCCTCCATTTTAAACAAACATTCGAGACCTTA 537
 QY 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
 |||||
 Db 538 TACAATACTTAGTTAAATAAATTCAGCATTTACTTAATTAACTTAAAGCAAGATTAAC 597
 QY 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
 |||||
 Db 598 GATTGTAATGTTGAAAGATGACGACATGTTAAATACTTAACCTTACTGATTTAAAA 657
 QY 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
 |||||
 Db 658 GCAATTTGATGACAAATAGATCTTTTAAAAACACTAACGACTTCGAAGCAATTTAAAAA 717
 QY 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
 |||||
 Db 718 TTGATAAATGATGATACGAAAGAAAGATATGCTTGGCAAAATTAATAGTACAGGATGAT 777
 QY 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
 |||||
 Db 778 CAAATTTTCTTAATACAAATATATACAAATTAATGAAGAAATTTCCAAGATATGTTA 837
 QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
 |||||
 Db 838 AACATTTTCAACACCAATGCGTAAACAAACAAATGTCCAGAAATTTCTGGATCTTCAGA 897
 QY 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGlyAspLys 377
 |||||
 Db 898 CATTTAGATCAAGAGAAAGAAATGTAATGTTTATTAAATTACAAACAAAGGTGATAAA 957
 QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397
 |||||
 Db 958 TGTGTTGAAATCCCAAGTCTACTTGTACGAAATAATGTTGGTGTGTCATGCAGATGCC 1017
 QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
 |||||
 Db 1018 ACATGTACCGAAGAGATTCAGGTAGCAGCAAGAAACAAATCACATGTGAATGTACTAAA 1077
 QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
 |||||
 Db 1078 CTGATCTTATCCACTTTTTCGATGGTATTCTTCGAGT 1116
 RESULT 26
 PFU20655 1116 bp DNA linear INV 14-FEB-1995
 LOCUS Plasmodium falciparum clone P27C merozoite surface protein 1
 DEFINITION

(MSP-1) gene, 42 kDa C-terminal region, partial cds.
 U20655
 U20655.1 GI:664847
 malarial parasite.
 Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 Shi.Y.-P., Alpers,M.P., Povoa,M.M., Nahlen,B.L., Dloo,A.G. and Lal,A.A.
 Sequence of the C-terminal region of merozoite surface protein-1 (MSP-1) in field-derived Plasmodium falciparum
 Unpublished
 Shi.Y.-P.
 Direct Submission
 Submitted (06-FEB-1995) Ya-Ping Shi, Molecular Vaccine Section,
 Bldg 22, Rm. 8F12, DPD/CDC 4770, Buford Highway, Chamblee, GA 30341
 Location/Qualifiers
 1..1116
 /organism="Plasmodium falciparum"
 /isolate="infected human blood sample from Papua New Guinea"
 /db_xref="taxon:5833"
 /clone="P27C"
 1..1116
 /gene="MSP-1"
 <1..>1116
 /gene="MSP-1"
 /codon_start=1
 /product="merozoite surface protein 1, 42 kDa C-terminal region"
 /protein_id="AAA62215.1"
 /db_xref="GI:664848"
 /translation="AISVTMDNILSGFENEYDVIYKPLAGVYRSUKKQIEKNITFFN
 LKNDILNSRLKKRYFDLVDLESLMQFKHSSNEYIIEESFKLNSSEKNTLLKSYK
 YIKESVNDIKFAQEGISYIEKVLAKYKDDLESIKKVEKEEFSSPTTPSPAKT
 DEQKESKFLPLNIETLYNNLVNKIDYILNLKAKINDCNVENEAKHVKTKLSDL
 KAIIDKIDLFKNTDFEAKIKLINDTKDKMLKLTGLVQNFNTITSKIEGKFO
 DMLNTSQHCVKQCPENSGCFRHIDERECKLLNYKQEGDKCVENPNPTCNEGG
 CDADATCTEEDSGSSRKKTCTCEKTPDSYPLFDGIFCS"
 BASE COUNT 486 a 146 c 160 g 324 t
 ORIGIN

Alignment Scores:
 Pred. No.: 6.87e-126 Length: 1116
 Score: 1917.50 Matches: 368
 Percent Similarity: 98.93% Conservatives: 1
 Best Local Similarity: 98.66% Mismatches: 3
 Query Match: 84.29% Indels: 1
 DB: 3 Gaps: 1
 US-10-057-531A-2 (1-431) x PFU20655 (1-1116)

QY 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
 |||||
 Db 1 GCAATATCTGTCCAAATGGATAATATCTCTCAGGATTTGAAATGAATATGATGTATA 60
 QY 78 TyrLeuLysProIleuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
 |||||
 Db 61 TATTTAAACCTTTTACTGGAGTATATAGAAGCTTTAAAAACAAATTTGAAAAACATT 120
 QY 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
 |||||
 Db 121 ATTACATTTAAATTTAAATTTGAACGATATCTTAAATTCAGTCTTAAGAACCGAAATAT 180
 QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
 |||||
 Db 181 TTCTTAGATGATTTAGAAATCTGATTTAATGCAATTTAAACATATATCTCTCAATGAATAC 240
 QY 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSer 157
 |||||
 Db 241 ATTATTGAAGATTCATTTAAATTTAATTTCAATTCAGAAACAAACACACTTTTAAAAAGT 300

QY 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
 Db 301 TACAATATATAAAGAATCATAGTAAATATTAATTTGCACAGAGGTATAAGT 360
 QY 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValle 197
 Db 361 TATTATGAAAGGTTTGTAGCGAAATATAAGGATGATTAGATCAATTAATAAAGTTATC 420
 QY 198 LysGluGluLysGluLysPheProSerProProThrThrProProSerProAlaLys 217
 Db 421 AAAGAGAAAGAGGAG---TTCCCATCATCACCACCAACACCTCCGTCCACGACAAA 477
 QY 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
 Db 478 ACAGACGAACAAAGAGAAAGTAAGTTCCTCCATTTTAAACAAACATTGAGACCTTA 537
 QY 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrIleuIleAsnLysAlaLysIleAsn 257
 Db 538 TACAATAACTTAGTTAATAAATTTGAGCATTTACTTAATTAATTAAGGCAAGATTAAAC 597
 QY 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
 Db 598 GATTGTAAATGTTGAAAGAAATGAGCACATGTTAAATAACTAACTTAGTGATTAAAA 657
 QY 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
 Db 658 GCAATTGATGACAAATAGATCTTTTAAAAAACAACACTAACGACTTCGAAGCAATTA 717
 QY 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuSerThrGlyLeuVal 317
 Db 718 TTGATAATGATGATACGAAAGAAAGATGCTTGGCAATTTACTTAGTACAGGATTAGT 777
 QY 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
 Db 778 CAAATTTTCTTANTACATAATATCAAAATTAATTAAGAGAAATTTCCAGATATGTTA 837
 QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
 Db 838 AACATTTCAACAACCAATGCGTAAAAAACAATGTCCAGAAAAATTTCTGGATGTTT 897
 QY 358 HisLeuAspGluAtgGluGluCysLysCysLeuLeuAsnTyrLysGlnGlyAspLys 377
 Db 898 CATTTAGATGAAGAGAGAAATGTAATGTTTATTAATTAACAACGAAGAGGTGATAA 957
 QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla 397
 Db 958 TGTGTGAAATCCAAATCCTACTTGTGAACGAAATAATGTTGATGATGTCAGATGCC 1017
 QY 398 ThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys 417
 Db 1018 ACATGTACCGAAGAGATTCAGGTAGCAGCAGAAAGAAAAATCACATGTGAATGTACTAA 1077
 QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
 Db 1078 CCTGATCTTATCCATTTTTCGATGGTATTTTTCGCACT 1116
 RESULT 27
 PFU20656 1116 bp DNA linear INV 14-FEB-1995
 LOCUS Plasmodium falciparum clone P27D merozoite surface protein 1
 DEFINITION (MSP-1) gene, 42 kDa C-terminal region, partial cds.
 U20656
 U20656.1 GI:664849
 SOURCE malarial parasite.
 ORGANISM Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 REFERENCE 1 (bases 1 to 1116)
 AUTHORS Shi,Y.-P., Alpers,M.P., Povoa,M.M., Nahlen,B.L., Oloo,A.G. and Lal,A.A.
 TITLE Sequence of the C-terminal region of merozoite surface protein-1 (MSP-1) in field-derived Plasmodium falciparum

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1116)
 AUTHORS Shi,Y.-P.
 TITLE Direct Submission
 JOURNAL Submitted (06-FEB-1995) Ya-Ping Shi, Molecular Vaccine Section,
 Bldg 22, Rm. 8F12, DPD/CDC 4770, Buford Highway, Chamblee, GA 30341
 FEATURES Location/Qualifiers
 source 1..1116
 /organism="Plasmodium falciparum"
 /isolate="infected human blood sample from Papua New Guinea"
 /db_xref="taxon:5833"
 /clone="P27D"
 1..1116
 /gene="MSP-1"
 <1..>1116
 /gene="MSP-1"
 /codon_start=1
 /product="merozoite surface protein 1, 42 kDa C-terminal region"
 /protein_id="AAA62216.1"
 /db_xref="GI:664850"
 /translation="ATSVTMDNIIISGFENEYDVIYLPKLAGVYRSLLKQIEKNITIFN
 LNUNDLNLSRLKKRKYFLDVLSDLKQFKHISSENYTIEDSFLLNSEQKNTLLKSYK
 YIKESVENDIKFAQEGISYERVLAKDDLESIKKVIKEKEFPSPPTPPSPAKT
 DEQKESKFLPFLTNIETIYNNLVNKIDDDYILNKKAKINDCNVNEARHVKITKLSDL
 KADDKRIDLFKNTNDFEAIKKLINDDTKDKMLKLLSTGLVQNFPTIISKLEGKFKQ
 DMLNISOHCVKKQCPNSGCFRHLDERECKLLNYKQEGDKCVENPNPTCENNING
 CDADATCTEEDSGSSRRKKTCTCECTPDSYFLFDGIFCS"
 BASE COUNT 486 a 146 c 160 g 324 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 6.87e-126 Length: 1116
 Score: 1917.50 Matches: 368
 Percent Similarity: 98.93% Conservativeness: 1
 Best Local Similarity: 98.66% Mismatches: 3
 Query Match: 84.29% Indels: 1
 DB: 3 Gaps: 1
 US-10-057-531A-2 (1-431) x PFU20656 (1-1116)
 QY 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValle 77
 Db 1 GCAATATCTGTCACATGGATTAATCTCTCAGGATTTGAAATGAATATGATGTATA 60
 QY 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
 Db 61 TATTTAAACCTTTAGCTGGAGTATATAGAACCTTAAAAAACAATTTGAAAAACATT 120
 QY 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
 Db 121 ATTACATTTAATTTAAATTTTCAACGATATCTTAAATTCACCTCTTAAGAACGAAATAT 180
 QY 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
 Db 181 TTCTTAGATGTATTAGATCTGATTTAATGCAATTTTAAACATATATCTCAATGAATAC 240
 QY 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSer 157
 Db 241 ATTTATGAAGATTCATTTAAATTTATGAAATTCAGAACAAAAAACACACTTTTAAAAAGT 300
 QY 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
 Db 301 TACAAATATATAAAGAAATCAGTAGAAAAATGATTTAAATTTGCACAGGAGGTATAAGT 360
 QY 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValle 197
 Db 361 TATTATGAAAGGTTTGTAGCGAAATATAAGGATGATTAGATCAATTAATAAAGTTATC 420
 QY 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
 Db 421 AAAGAGAAAGGAG---TTCCCATCATCACCACCAACACCTCCGTCCACGACAAA 477

Qy	218	ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu	237
Db	478	ACAGACGAACAAGAAAGAAAGTAAGTCCCTTCATTTTACAAACATTTAGACCTTA	537
Qy	238	TyrAsnAsnLeuValAsnLysLysLysLeuAspTyrLeuIleAsnLeuLysLysLysLys	257
Db	538	TACAATAACTAGTTAATAAATGACGATTACTTAACTTAACTTAAAGCAAGATTAAAC	597
Qy	258	AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys	277
Db	598	GATTGTAATGTTGAAAAAATGAGCACATGTTAAATACTAACTTAGTGNATTTAAA	657
Qy	278	AlaIleAspAspLysLysLysLeuPheLysAsnProTyrAspPheGluAlaIleLysLys	297
Db	658	GCAATTGATGACAAAATAGATCTTTTAAAAACACTAACGACTTCGAAGCAATTTAAA	717
Qy	298	LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal	317
Db	718	TTGATAAATGATGATACGAAAAAAGATATGCTTGGCAATTTACTTAGTACAGGATTAGT	777
Qy	318	GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu	337
Db	778	CAAAATTTTCCTAATACATATATCAAAATTAATTGAAGGAAAAATTCGAAGATATGTA	837
Qy	338	AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg	357
Db	838	AACATTTTCAACACCAATCGCTAAAAAACAATGTCCAGAAATTTCTGGATGTTTCAGA	897
Qy	358	HisLeuAspGluArgGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys	377
Db	898	CATTTTAGATGAAGAGAAGAATGTAATGTTTATTAATACAAACAAGAGGTGATAA	957
Qy	378	CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla	397
Db	958	TGTGTTGAAATCCAAATCCTACTTGTACGAAATAATGTTGGATGTCATCGAGATGCC	1017
Qy	398	ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys	417
Db	1018	ACATGTACCGAAGAGATTCAGGTAGCAGCAGCAAGAAAAATCACATGTGAATGTACTAA	1077
Qy	418	ProAspSerTyrProLeuPheAspGlyIlePheCysSer	430
Db	1078	CCTGATTCATCCACTTTTTCGATGTTATTTCTGCGAGT	1116
RESULT	28		
PFMEZSALD		1956 bp	DNA linear
LOCUS			INV 22-SEP-1995
DEFINITION			P. falciparum (RO-71) gene for merozoite surface antigen 1.
ACCESSION			235329
VERSION			235329.1
KEYWORDS			merozoite surface antigen 1 gene.
SOURCE			Plasmodium falciparum.
ORGANISM			Plasmodium falciparum
REFERENCE			Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
AUTHORS			Tolle, R., Bujard, H. and Cooper, J.A.
TITLE			Plasmodium falciparum: Recombination within the C-terminal region of merozoite surface antigen-1
JOURNAL			unpublished
REFERENCE			2 (bases 1 to 1956)
AUTHORS			Olafsson, P., Matile, H. and Certa, U.
TITLE			Plasmodium falciparum: the repetitive MSA-1 surface protein of the RO-71 isolate is recognized by mouse antibody against the nonrepetitive repeat block of RO-33
JOURNAL			Exp. Parasitol. 74 (4), 381-389 (1992)
MEDLINE			92275047
PUBMED			1592091
REFERENCE			3 (bases 1 to 1956)
AUTHORS			Tolle, R.
TITLE			Direct Submission
JOURNAL			Submitted (18-JUL-1994) Ralf Tolle, Abt. Prof. H. Bujard, Zentrum fur Molekulare Biologie, (ZMBH), Univ. Heidelberg, Im Neuenheimer
Qy	218	ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu	237
Db	478	ACAGACGAACAAGAAAGAAAGTAAGTCCCTTCATTTTACAAACATTTAGACCTTA	537
Qy	238	TyrAsnAsnLeuValAsnLysLysLysLeuAspTyrLeuIleAsnLeuLysLysLys	257
Db	538	TACAATAACTAGTTAATAAATGACGATTACTTAACTTAACTTAAAGCAAGATTAAAC	597
Qy	258	AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys	277
Db	598	GATTGTAATGTTGAAAAAATGAGCACATGTTAAATACTAACTTAGTGNATTTAAA	657
Qy	278	AlaIleAspAspLysLysLysLeuPheLysAsnProTyrAspPheGluAlaIleLysLys	297
Db	658	GCAATTGATGACAAAATAGATCTTTTAAAAACACTAACGACTTCGAAGCAATTTAAA	717
Qy	298	LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal	317
Db	718	TTGATAAATGATGATACGAAAAAAGATATGCTTGGCAATTTACTTAGTACAGGATTAGT	777
Qy	318	GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu	337
Db	778	CAAAATTTTCCTAATACATATATCAAAATTAATTGAAGGAAAAATTCGAAGATATGTA	837
Qy	338	AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg	357
Db	838	AACATTTTCAACACCAATCGCTAAAAAACAATGTCCAGAAATTTCTGGATGTTTCAGA	897
Qy	358	HisLeuAspGluArgGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys	377
Db	898	CATTTTAGATGAAGAGAAGAATGTAATGTTTATTAATACAAACAAGAGGTGATAA	957
Qy	378	CysValGluAsnProAsnProThrCysAsnGluAsnGlyGlyCysAspAlaAspAla	397
Db	958	TGTGTTGAAATCCAAATCCTACTTGTACGAAATAATGTTGGATGTCATCGAGATGCC	1017
Qy	398	ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys	417
Db	1018	ACATGTACCGAAGAGATTCAGGTAGCAGCAGCAGCAAGAAAAATCACATGTGAATGTACTAA	1077
Qy	418	ProAspSerTyrProLeuPheAspGlyIlePheCysSer	430
Db	1078	CCTGATTCATCCACTTTTTCGATGTTATTTCTGCGAGT	1116
RESULT	28		
PFMEZSALD		1956 bp	DNA linear
LOCUS			INV 22-SEP-1995
DEFINITION			P. falciparum (RO-71) gene for merozoite surface antigen 1.
ACCESSION			235329
VERSION			235329.1
KEYWORDS			merozoite surface antigen 1 gene.
SOURCE			Plasmodium falciparum.
ORGANISM			Plasmodium falciparum
REFERENCE			Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
AUTHORS			Tolle, R., Bujard, H. and Cooper, J.A.
TITLE			Plasmodium falciparum: Recombination within the C-terminal region of merozoite surface antigen-1
JOURNAL			unpublished
REFERENCE			2 (bases 1 to 1956)
AUTHORS			Olafsson, P., Matile, H. and Certa, U.
TITLE			Plasmodium falciparum: the repetitive MSA-1 surface protein of the RO-71 isolate is recognized by mouse antibody against the nonrepetitive repeat block of RO-33
JOURNAL			Exp. Parasitol. 74 (4), 381-389 (1992)
MEDLINE			92275047
PUBMED			1592091
REFERENCE			3 (bases 1 to 1956)
AUTHORS			Tolle, R.
TITLE			Direct Submission
JOURNAL			Submitted (18-JUL-1994) Ralf Tolle, Abt. Prof. H. Bujard, Zentrum fur Molekulare Biologie, (ZMBH), Univ. Heidelberg, Im Neuenheimer

REFERENCE	Feld 282, Heidelberg, 69120, Germany
AUTHORS	4 (bases 1 to 1956)
TITLE	Tolle,R., Bujard,H. and Cooper,J.A. Plasmodium falciparum: variations within the C-terminal region of merozoite surface antigen-1
JOURNAL	Exp. Parasitol. 81 (1), 47-54 (1995)
MEDLINE	95354793
PUBMED	7628566
FEATURES	
source	Location/Qualifiers 1..1956 /organism="plasmodium falciparum" /isolate="RO-71" /db_xref="taxon:5833" /chromosome="9" /dev_stage="blood stage" 1..1956 /gene="gp190, MSA1, MSP1, PMWSA" <1..1956 /gene="gp190, MSA1, MSP1, PMWSA" /codon_start=1 /product="merozoite surface antigen 1 gene, 3+-end" /protein_id="CAA84558.1" /db_xref="GI:535258" /db_xref="SPTREMBL:Q25924" /translation="FNKKELGQDKMQIKKLLKLEQESKLSLNPNVYNLFQSFV FNKKKEAETAEITENTLENTILLKHYKGLVYKNGESSPKLTLSVSIQTDENYANLE KFRALSKIDGKLNLDNLHLGKKLSSGLHLLITELKEVINKNKYNSTNSNNKKV NEALSYENFLPEAKVTVPDPTVPSPLSVRSVSGSGSTKEETOIPTSGLLLE LQOQVQLQNYDEEDDSLVLVPIFGESENDDEYLDQVVTGEALSVTMDNLISGFENEID VYILKPLAGVYRSLLKQIEKNIFTNLNLDNLSRLKKRYFLDVLSDLMQFKHIS SNEYIIESDKLLNSEQNTLLSKYIKESVENDIKFAQEGISYKYKLVAKYKDDLE STKKYIKKEEPSPPTPPSPAKTDEQKESKFLPTNIETLYNNLVAKYKDDLE NLKAKINDCNVEKDEAHVKITKLSDLKATDDKIDLPKHNDFEAIKKLNDOTKKDML GKLLSTGLVQNPNTIISKLESGKQDMQNLINISOHOCVKKQCPQNSGCPRHLDERECK CLLVKQEGDKCVENPNTCENNGCCADACTBEDSGSNKKITCECTKPDSPPLP DGIFVCSNSFLIGISFLILLMLILKSF1"
BASE COUNT	855 a 263 c 265 g 573 t
ORIGIN	
Alignment Scores:	
Pred. No.:	1..34e-125 Length: 1956
Score:	1917.00 Matches: 370
Percent Similarity:	96.19% Conservative: 9
Best Local Similarity:	93.91% Mismatches: 9
Query Match:	84.26% Indels: 6
DB:	3 Gaps: 2
US-10-057-531A-2 (1-431) x PFMEZSALD (1-1956)	
Qy	38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57 : ::

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
 Db 1075 TACAATATATAAAGAAATCATAGTAAATGATATTAATTTGCACAGGAAGGTATAAGT 1134
 Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle 197
 Db 1135 TATTATGAAGAGTTTGTAGCCAAATATAAGGATGATTTAGAAATCAATTAATAAAGTATTATC 1194
 Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
 Db 1195 AAGAAGAAAGAGGAG---TTCCCATCATCACCACCAACACACCTCCGTACACGACAAA 1251
 Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
 Db 1252 ACAGACGACAAAAGAGAAAGTAAAGTCTCTCCATTTTAAACAAACATTGAGACCTTA 1311
 Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
 Db 1312 TACAATAACTTAGTTAATAAATTGACGATTACTTAATTAATTAATTAAGGCAAGATTAAAC 1371
 Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
 Db 1372 GATTGTAATGTTGAAAGATGAAGCACATGTTAAATAAATACTTAACTAGTATTAAAA 1431
 Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
 Db 1432 GCAATTGATGACAAAATAGATCTTTTAAACCAATTAACGATCTCGAAGCAATTAACAAA 1491
 Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
 Db 1492 TTGATAATGATGATACGAAAAGATATGCTTGCGAAATTAATTAATTAATTAATTAATTA 1551
 Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
 Db 1552 CAAAATTTCTTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1611
 Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
 Db 1612 AACATTTTCAACACCAATCGTAAAAAACAATGTCACCAAAATCTCGGATGTTTCAGA 1671
 Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
 Db 1672 CATTAGATGAAGAGAGATGTAATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTA 1731
 Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
 Db 1732 TGTGTTGAAAATCCAAATCTTACTTGAACGAAAATAATGTTGATGTCAGATGCC 1791
 Qy 398 ThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys 417
 Db 1792 AAATGTACCGAAGAGATTCAGGTAGCAACGGAAGAAAATACATGTGAATGTACTAAA 1851
 Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
 Db 1852 CCTGATCTTATCCACITTTTCGATGGTATTTTCTGCAGTTCC 1893

RESULT 29
 PFAMSP815
 LOCUS P. falciparum DNA 1726 bp DNA linear INV 29-MAY-2002
 DEFINITION P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.
 ACCESSION D13346
 VERSION D13346.1 GI:391792
 KEYWORDS EGF-like domains; major merozoite surface protein precursor.
 SOURCE Plasmodium falciparum (isolate 815) merozoite, DNA.
 ORGANISM Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 1726)
 JONGWUTIWES.S., TANABE,K. and KANBARA,H.
 AUTHORS Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates
 TITLE falciparum from field isolates
 JOURNAL Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)

93295445
 2 (bases 1 to 1726)
 JONGWUTIWES.S.
 TITLE Direct Submission
 JOURNAL Submitted (05-OCT-1992) Somchai Jongwutiwes, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)
 COMMENT Submitted (05-OCT-1992) to DDBJ by:
 Somchai Jongwutiwes
 Department of Protozoology
 Institute of Tropical Medicine
 Nagasaki University
 12-4 Sakamoto-machi
 Nagasaki 852
 Japan
 Phone: 0958-47-2111x3747
 Fax: 0958-47-6607
 FEATURES
 source
 1. .1726
 /organism="Plasmodium falciparum"
 /db_xref="taxon:5833"
 /chromosome="9"
 /dev_stage="merozoite"
 1. .1712
 /gene="MSP1"
 <1. .1712
 /gene="MSP1"
 /codon_start=3
 /product="major merozoite surface protein precursor"
 /protein_id="BAA02607.1"
 /db_xref="GI:391793"
 /translation="PLKTLSEVSIQTDNYANLEKFRVLSKIDGKLNLDNHLGKKKLS
 FLSSGLHLITELKEVINKNTGNSPENNKKVNEALKSYENFLPEAKYTVTVTPPQ
 PDVTPSLSVRVSGSGSTKEETQIPTSGSLTLEQOVVOLQNVDEEDSLVLPFG
 ESDNDYLDQVVTGEAISVTMDNILLSGFENEYDVLYKPLAGVRSKKQIEKNIT
 FNUNLDILNRLKKRKYFLDLESLLMQFKHSSNEYIEDSFKLNSQKNTLLKS
 YRKIESVENDIKFAQEGISYEVLYAKYKDDLESKKVKEEKEFPSSPTTPPSA
 KTDQKESKFLPLTNIETLYNNLVNKIDYILNKAKINDCNVEKDEAHVKILS
 DLKAIIDKIDLPKNTDFAIKKLIINDTKKMLGLSLTGLVQNPNTTISLIEBK
 FQMLNLSHQCVKQCPNSGCFRHLDERECKLLNLYKQEGDKCVENPNPTCNENN
 GGCDAKACTEEDSGSNGKKTCECTKPDSPFLFDGIFCSSSNFLGTSFLILMLILY
 SFI"
 BASE COUNT 741 a 232 c 245 g 508 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,62e-125 Length: 1726
 Score: 1915.00 Matches: 370
 Percent Similarity: 95.94% Conservatives: 8
 Best Local Similarity: 93.91% Mismatches: 10
 Query Match: 84.18% Indels: 6
 DB: 3 Gaps: 2
 US-10-057-531A-2 (1-431) x PFAMSP815 (1-1726)
 Qy 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57
 Db 486 TCCGAAGATAATGACGAATATTAGATCAAGTAGTAAGTGGAGAA----- 530
 Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
 Db 531 GCAATATCTGCACAAATGGGAATATCTCTCAGGATTTGAAAATGAATATGATGTATA 590
 Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
 Db 591 TATTAAACCTTTAGCTGGAGTATATAGAAGCTTAAAAACAACAAATGAAAAACATT 650
 Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
 Db 651 ATTACATTTAATTTAAATTTGAACGATATCTTAAATTCACGCTCTTAAGAAACGAAATAT 710
 Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137

Db 591 TATTATAAACCTTTAGCTGGAGTATATAGAACCTTAAAAAACAATAATCAAAAAACATT 650
 Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
 Db 651 ATTACATTTAAATTTAAATTTGAACGATATCTTAAATTCACGCTCTTAAGAAACGAAATAT 710
 Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
 Db 711 TTCTTAGATGATTAGATCTGATTATGCAATTTAATCAATATATCTCAATCAATGATAC 770
 Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
 Db 771 ATTATGAGATTCATTTAAATTTATGATTCAGAACAAAAACACACTTTTTAAAGAT 830
 Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
 Db 831 TACAATATATAAAGAAATCAGTACAAATGATATTAATTTGCACAGGATATAAGT 890
 Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle 197
 Db 891 TATTATGAAAGGTTTACGGAATATATAGGATGATTTAGATCAATTAATAAAGTTATC 950
 Qy 198 LysGluGluLysGluLysPheProSerProProThrThrProProSerProAlaLys 217
 Db 951 AAGAAGAAAAAGGAG--TTCCCATCATCACCACCAACAACACCTCCGTCACCACAAAA 1007
 Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
 Db 1008 ACAGACGAAACAAAGAAAGAAAGTAAGTTCCTTCCTTCATTTTAAACAAACATTCGAGACCTTA 1067
 Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
 Db 1068 TACAATACTAGTTAATAAATTCACCATCTTAATAATTAATTAAGCAAGATTAAC 1127
 Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
 Db 1128 GATTCTAATGTTGAAAAAGATGAACACACATGTTAAATAAATACTAACTTAGTGATTANA 1187
 Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
 Db 1188 GCAATTGATGACAAAATAGATCTTTTAAAAACACATCAACGACTTCGAAGCAATTAATAA 1247
 Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGluLysLeuLysLeuSerThrGlyLeuVal 317
 Db 1248 TTGATAAATGATGATACGAAAAAGATATGCTTGCAATTAATCTTAGTACAGGATAGTT 1307
 Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
 Db 1308 CAAAAATTTCTCTAATAACAATAATCAAAATTAATTTGAAGGAAAAATTCAGATATGTA 1367
 Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
 Db 1368 AACATTTCCACACACCAATGCGTAAAAAACAATGTCAGAAAAATCTCGATGTTTCAGA 1427
 Qy 358 HisLeuAspGluArgGluGluCysLysCysLysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
 Db 1428 CATTTAGATGAAGAGAGAGATGTAATGTTTATTAATTAACAACAAGAGTGATATA 1487
 Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
 Db 1488 TGTGTTGAAATCCAAATCCTACTTGTAAACGAAAAATAATGTTGGATGATGTCAGATGCC 1547
 Qy 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417
 Db 1548 AAATGTCGGAAGAGATTCAGGTAGCAACGAAAGAAAAATCACATGTGAATGTACTAAA 1607
 Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
 Db 1608 CTGATTCTTATCCACTTTTCGATGGTATTTCTGTCAGTTCC 1649

RESULT 31

PFAMSP8352

LOCUS

1726 bp

DNA

linear

INV

29-MAY-2002

DEFINITION

P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.

ACCESSION

D13349

VERSION

1 GI:391804

KEYWORDS

EGF-like domains; major merozoite surface protein precursor.

SOURCE

Plasmodium falciparum

ORGANISM

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE

1 (bases 1 to 1726)

AUTHORS

Jongwutives, S., Tanabe, K. and Kanbara, H.

TITLE

Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates

JOURNAL

Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)

REFERENCE

2 (bases 1 to 1726)

AUTHORS

Jongwutives, S.

TITLE

Direct Submission

JOURNAL

Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747)), Fax:0958-47-6607)

COMMENT

Submitted (05-OCT-1992) to DDBJ by:

Somchai Jongwutives

Department of Protozoology

Institute of Tropical Medicine

Nagasaki University

12-4 Sakamoto-machi

Nagasaki 852

Japan

Phone: 0958-47-2111x3747

Fax: 0958-47-6607

FEATURES

source

Location/Qualifiers

1..1726

/organism="Plasmodium falciparum"

/db_xref="taxon:5833"

/chromosome="9"

/dev_stage="merozoite"

1..1712

/gene="MSP1"

<1..1712

/gene="MSP1"

/codon_start=3

/product="major merozoite surface protein precursor"

/protein_id="BAA02610.1"

/db_xref="GI:391805"

/translation="PLKTVSEVSIQTEDNYANLEKFRVLSKIDGKLNLDNLHGKKKLS

FLSSGLHLITELAEVKNKNTGNSPSENKKNYNEALKSVENLPKAKVTVTTPQ

PDVTPSPLSVRVSGSGSTKEETQIPTSGSLTTELQOVQSONYDEEDDSLVLPIFG

ESENDEYLDQVVTGEALSVTMDNLGSENEYDVIYLLKPLAGVIRSLUKKIEKNII

FNLNDILNSRLAKRKYFLDVLSDLMQFKHISSEYIIEDESKLNSQKNTLLKS

YKYIKESVNDIKFAQEGISYKVLAKYKDDLESIKKVIKEKEFPSPPTTPSPA

KTDQKKESKPLPFTNIETIYLNVLNKIDDIYLNKAKINDCNVEKDEAHVITKLS

DLKAIIDDKIDLFKNTDPEAIKKLINDTKKMLGKLLSTGLVNFPTIISKIEGK

FQDLNLSQHCYKVKQCPNSGCFRHLDERECKLLNKGKGCVCNPNPTCNENN

GGCDANAKTEEDSGSGNKKITCCTKPDSTPLFDGIFGFCSSNSFLGSLFLMLILY

SFI"

BASE COUNT 741 a 233 c 245 g 507 t

ORIGIN

Alignment Scores:

Pred. No.: 1.62e-125 Length: 1726

Score: 1915.00 Matches: 370

Percent Similarity: 95.94% Conservativeness: 8

Best Local Similarity: 93.91% Mismatches: 10

Query Match: 84.18% Indels: 6

DB: 3 Gaps: 2

US-10-057-531A-2 (1-431) x PFAMSP8352 (1-1726)

Qy 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57

Db 486 TCCGAGATATGACGATATTTAGATCAAGTAGTAAGTGGAGAA-----530

[illegible]

DB:	3	Gaps:	2
US-10-057-531a-2 (1-431) x PFAMSP838 (1-1726)			
Qy	38	ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet	57
Db	486	TCCGAAGATAATGACGAATATTTAGATCAAGATGAATGACTGGAGAA-----	530
Qy	58	AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle	77
Db	531	GCAATATCTGTCACATGGATTAATCTCTCAGGATTTGAAATGAATGATGATTATA	590
Qy	78	TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle	97
Db	591	TATTAAACCTTTAGCTGGAGTATATAGAAGCTTAAAAAACAATTTGAAAAAACA	650
Qy	98	PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr	117
Db	651	ATTACATTTAATTTAAGCATATCTTAATTCACGCTCTTAAGAACGAAATAT	710
Qy	118	PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr	137
Db	711	TTCTTAGATGATTAGAATCTGATTAATGCAATTTAAACATATATCTCAATGAATAC	770
Qy	138	IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer	157
Db	771	ATTATTGAAGATTCATTTAAATTTATTGAATTCAGAACAAAAAACACACTTTTAA	830
Qy	158	TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer	177
Db	831	TACAAATATATAAAGAAATCAGTAGAATAATGATTAATTTGCACAGGAAGGTATAGT	890
Qy	178	TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle	197
Db	891	TATTATGAAGAGTTTAGCGAAATATAAGGATGATTAGATCAATTAATAAAGTATC	950
Qy	198	LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys	217
Db	951	AAAAAGAAAAAGGAG---TTTCCATCATCATCACCAACAACACCTCGCTCACGACAAA	1007
Qy	218	ThrAspGluGlnLysLysLysPheLysPheLeuProPheLeuThrAsnIleGluThrLeu	237
Db	1008	ACACACGACAAAGAGAGAAAGTAAGTTCTCTCCATTTTAAACAACATTTGACACCTTA	1067
Qy	238	TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn	257
Db	1068	TACAATAACTTAGTTAATAAAATTCAGCATTTACTTAATTAATTAAGGCAACATTAAC	1127
Qy	258	AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys	277
Db	1128	GATTGTAATGTTGAAAAGATGAAGCACATGTTAAAAATACTAAACTTAGTATTAAAA	1187
Qy	278	AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys	297
Db	1188	GCAATGATGACAAATAGATCTTTTAAAAAACAATACAGCTTCGAGCAATTAATAAAA	1247
Qy	298	LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal	317
Db	1248	TTGATAATGATGATACGAAAAAGATGCTTGCGAAATTTACTTAGTACAGGATTAGTT	1307
Qy	318	GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu	337
Db	1308	CAAAATTTTCTTAATACAAATATATCAAAATTAATTAAGGAAAAATTCCAAGATATGTTA	1367
Qy	338	AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg	357
Db	1368	AACTTTTCAACCAACCAATGCGTAAAAAACAATGTCAGAAAAATCTCGAGTTTTCAGA	1427
Qy	358	HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys	377
Db	1428	CATTTAGATGAAGAGAGAAATGTAATGTTTATTAAATTAACAACAAGAGGTGATAA	1487
Qy	378	CysValGluAsnProAsnProThrCysAsnGluAsnGlnGlyCysAspAlaAspAla	397

Db	1488	TGTGTTGAAAATCCAAATCCTACTTGTAAAGAAAATAATGTTGATGTCAGATGCC	1547
Qy	398	ThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys	417
Db	1548	AAATGTACCGAAGAGATTCAGTAGCAGCAAGAAAGAAATTCACATGTGAATGACTAAA	1607
Qy	418	ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431	
Db	1608	CCTGATCTTATCCACTTTTCGATGGTATTTCCTGCAAGTTCC	1649
RESULT 33			
PFAMSP842			
LOCUS	P. falciparum DNA for the precursor to the major merozoite surface	1726 bp	DNA linear INV 29-MAY-2002
DEFINITION	proteins, C-terminal.		
ACCESSION	D13352		
VERSION	D13352.1	GI:391812	
KEYWORDS	EGF-like domains; major merozoite surface protein precursor.		
SOURCE	Plasmodium falciparum (isolate 842) merozoite, DNA.		
ORGANISM	Plasmodium falciparum		
REFERENCE	1 (bases 1 to 1726)		
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
TITLE	Jongwutives,S., Tanabe,K. and Kanbara,H.		
JOURNAL	Sequence conservation in the C-terminal part of the precursor to		
MEDLINE	the major merozoite surface proteins (MSP1) of Plasmodium		
REFERENCE	falciparum from field isolates		
AUTHORS	Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)		
JOURNAL	93295445		
MEDLINE	2 (bases 1 to 1726)		
REFERENCE	Jongwutives,S.		
AUTHORS	Direct Submission		
JOURNAL	Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical		
COMMENT	Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4		
	Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747),		
	Fax:0958-47-6607)		
	Submitted (05-OCT-1992) to DBJ by:		
	Somchai Jongwutives		
	Department of Protozoology		
	Institute of Tropical Medicine		
	Nagasaki University		
	12-4 Sakamoto-machi		
	Nagasaki 852		
FEATURES	source		
	0958-47-2111x3747		
	0958-47-6607.		
	Location/Qualifiers		
	1..1726		
	/organism="Plasmodium falciparum"		
	/db_xref="taxon:5833"		
	/chromosome="9"		
	/dev_stage="merozoite"		
	1..1712		
	/gene="MSP1"		
	<1..1712		
	/gene="MSP1"		
	/codon_start=3		
	/product="major merozoite surface protein precursor"		
	/protein_id="BAA02613.1"		
	/db_xref="GI:391813"		
	/translation="PLRTUSEVSIQTDNTVANLEKFRALSKIDGKLNLDNHLGKKKLS		
	FLSSGLHLITELKVEYKNNKNTGSPSENKKVNEALKSYENFLPEAKVTVVTPPO		
	PDTPTSPSVRSVSGSGSTKEETQIPTSGSLLTQLQVVOHQYQNBEDSDSLVLPFG		
	ESNDDEYLDQVVTGAI SVTMDNILSGFENEYDVTVLRPLAGVRSKVKQIKENIT		
	FNLNDILNSRLKRRKFLDVLDSLMQPKHHSNEYIIEDSFKLLNSPKNTLLKS		
	YKIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEKEPPSPPTTPPSA		
	KTDQKKESKFLPLFTNIETFLYNNLVNKIDDYILNLAKINDCNVKNDEAHVKIKLS		
	DLKAIIDKIDLFKNTDFEAIKKLINDTKMDLGLKLLSTGLYQNPNTIISKLEK		
	FQMLNATSQHCVKQCPNSGCFRLHDERECKLLNLYKQEGDKCPNPENPNENN		
	GGCDADAKCTEEDSGSGKKTCTCECTKPDSPYPLFDGIFCSCSSNFLGTSFLILMLILY		
	SFI"		
BASE COUNT	741 a 233 c 245 g 507 t		
ORIGIN			

Alignment Scores:

```

Pred. No.:      1.62e-125      Length:      1726
Score:          1915.00      Matches:      370
Percent Similarity: 95.94%      Conservative: 8
Best Local Similarity: 93.91%      Mismatches: 10
Query Match:      84.18%      Indels:      6
DB:              3          Gaps:      2

US-10-057-531a-2 (1-431) x PFAMSP842 (1-1726)

Qy 38 ThrAspAspAspAspAlaMetAlaSpIleGlySerIleGluGlyArgGlyThrMet 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 486 TCCGAAGATAATGACCAATAATTAGATCAAGTAGTAAGTGGAGAA-----530

Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 531 GCATATCTCTCACAAATGGAATAATCTCTCAGGATTTGAAANATGAATATGATTATA 590

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 591 TATTTAAACCTTTAGCTGGAGTATATAGAACCTTAAAAAACAATAATTCAAAAACATT 650

Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 651 ATTACATTTTAATTTAAATTTGAACGATATCTTTAAATTTACGCTTTAAGAAACGAAATAT 710

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 711 TTCCTTAGATGATTAGAAATCTGATTTAATGCAATTTAAACATATATCTCAATGAATAC 770

Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 771 ATTATTGAAGATTCATTTAAATTTATTGAATTCAGAACAAAAAACACACACTTTTAAAAAGT 830

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluLysSer 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 831 TACAATATATAAAAGAAATCAGTAGAAATGATATAATTTGACAGGAAGGTATAAGT 890

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValIle 197
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 891 TATTATGAAGAGCTTTACGGAATATAGGATGATTAGATCAATTAATAAAGTTATC 950

Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 951 AAAGAAGAAAAGGAG---TTCCCATCATCACCACCAACAACACCTCCGTCACCAGCAAAA 1007

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1008 ACAGACGAAACAAGAAAGAAAGTAGTTCTTCCATTTTAAACAACATTCAGACCTTA 1067

Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1068 TACAATACTTAGTTAATAAATTCAGGATTAATTAATTAATTAAGCAAGATTATAC 1127

Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1128 GATTCGTAATGTTGAAAGAAATGAAGCACATGTTAAATAACTAAACTTAGTGAATTTAAA 1187

Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1188 GCATTTGATGACAAATATAGATCTTTTAAACAACACTAACGACTTCGAGCAATTAATAA 1247

Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1248 TTGATAATGATGATACGAAAAAGATATGCTTGGCAATTAATTAATTAAGCAAGATTAGT 1307

Qy 318 GlnAsnPheProAsnThrIleLeuSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1308 CAAAAATTTTCCATAATAATAATATCAAAATTAATTAAGGAAAAATTCGAAGATATGTTA 1367

Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 1368 AACATTTCCAAACCAACCAATGCGTAAACCAAAACCAATGTCCAGAAAAATCTCGATGTTTCAGA 1427
Qy 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1428 CATTTAGATGAAGAAGAGAATGTAATGTTTATTAAATTTACAAACGAAGAGTGATAAA 1487
Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlnGlyCysAspAlaAspAla 397
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1488 TCTGTTGAAATCCAAATCCTACTTGTAAACGAAATAATGTTGGATGTGATCGAGATGCC 1547
Qy 398 ThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys 417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1548 AAATGTACCGAAGAAGATTGAGTAGCAACGGAAGAAAAATCACATGTGAATGTACTAAA 1607
Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1608 CCTGATCTTATCCACTTTTTCGATGATGTTTCTTCGAGTTCC 1649

RESULT 34
LOCUS AF218248 5085 bp DNA linear INV 24-JAN-2000
DEFINITION Plasmodium falciparum major merozoite surface antigen (gp195) gene, complete cds.
ACCESSION AF218248
VERSION AF218248.1 GI:6740099
KEYWORDS Plasmodium falciparum.
SOURCE Plasmodium falciparum.
ORGANISM Eukaryota; Alveolata, Apicomplexa; Haemosporidia; Plasmodium.
REFERENCE 1 (bases 1 to 5085)
AUTHORS Shan, Z.X., Yu, X.B., Li, X.R., Ma, C.L. and Fang, J.M.
TITLE Molecular cloning and sequence analysis of major merozoite surface antigen (gp195) gene of Plasmodium falciparum isolate FCC1/HN
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 5085)
AUTHORS Shan, Z.X.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Parasitology, Sun Yat-sen University of Medical Sciences, 74 Zhongshan Road II, Guangzhou, Guangdong 510089, P.R.China
FEATURES
     source
       1..5085
         /organism="Plasmodium falciparum"
         /isolate="FCC1/HN"
         /db_xref="taxon:5833"
         /country="China"
       1..5085
         /gene="gp195"
         /gene="gp195"
         /codon_start=1
         /product="major merozoite surface antigen"
         /protein_id="AAF27526.1"
         /db_xref="GI:6740100"
         /translation="MKIIFFLCSFLFFIINTOCVTHESQVELVKKLEALEDAVLGYS
LFQKEMVLNCGTSGTAVTSTPGSGSVTSGSGSVASVASGSGGSGVASGSGGSGNS
RTNPDSNDSDAKSYADLKHVRQNLFTIKELYPFLDLTNHMLTCDNHLKIFKY
LIDGVEINELLYKLFYDILLARLNDANCANSYQIPENLKRANLELDVKKIVFGY
RKPLDNKDNVGMEDYTKKNTTIANINELTSGSKTIDONKNADNEEGKKLYQAO
NLFYTNKLOEAHNLISVLEKRIDTLKKNKIKLLEDIDIKTDIAEKLTTGSKPNP
LPENKKKEVEGHEEKIKELAKTIKENIDSLFTDPLELEYLREKKNKVDVTPKSQDPT
KSQIPKVPYNGIYPLPLTDIHNSLAADNKNSTGDLMPDTEKIEKIKITDNKE
RKIFINNTIKQIDLEENINHTKEQNKLLLEDYKSKDYELLEKFEYEMKFNPNFDK
DVVDIFSGARTYINVEKQNNKFSNNSSNVYNOKLKALSYLSDYSLRKIGTSEKDF
NHYYLKTGLEDADIKLLEIKSENKILEKFKGLTHSANASLEYSIDVILQVOKVL
LKTKTEDLRKIELFKNAQLKDSIHVNLYKPNKPEPYLLIVLKEVDKLEKFPKV
KDLKKQAVLSSITQPLVAASETTEDGGHSTHTLSQSGETEETETVETGHTVITL
PPTQSPPREKVENSIENHSNDNSQALTUUKLDEFLUTKSTYCHYILVSNSS
MDQKLEVINLTPPEKNEKCDPLDLFIQNINIPAMYSLYDSMNNDLQHLFFELYQ
KEMIYVHLKKEENHKLLBEQKQITGTSSTSSPGNTVTNTAQSNATHQSNQOONA
SSTNTQNGVAVSSGPAVPESHDPLTVLSISNDLKGIVSLLNGLNKKTKVPNPLTISTT
ENEKFEYENLLKNDYTFNDIDIKQFVKNSKVITGLTETOKNALNENPHNKLQTLQSP
DLYNKYTKDLRLFNKKLGQDKMQIKKLTLLKEQLSKSLNSLNPHNQLQNSVFF

```


Alignment Scores:
Pred. No.: 5, 04e-125 Length: 5085
Score: 1915.00 Matches: 370
Percent Similarity: 95.94% Conservative: 8
Best Local Similarity: 93.91% Mismatches: 10
Query Match: 84.18% Indels: 6
DB: 3 Gaps: 2

US-10-057-531A-2 (1-431) x AF218248 (1-5085)

BASE COUNT 2244 a 671 c 678 g 1492 t
ORIGIN

KKKEAEIAETENTLENTKILLKHLKGLVYKNGESSPLKTLSEVSIQTDNYANLEK
FRVLSIDGKLNLDNLHGLKKFLSSGLHHLTELKVIKNNKNTYGNPSSENKKVN
EALKSYNEPEAKVTVTPPOPDPSPPLSVRSVSGSGSTKEETQIPTSGSLTEL
QYVQSONYDEEDSDVLWLP IGESEDNDEYLDVOVTEAISVMTDNILSGPENEYDV
IYLPLAGVRSLLKQIKENILITFNLANLDILNRLKKFKYFLDVLDESLOMOPKHLS
NEYIIEDESKLLNSEQKNLLKSKYKIKESVENDIKFAQGGSYEKVLAKYKDDLES
IKKVIKEEFSSPTTPPPSPATDEQKESKFLPFLNIETLYNNLKNKIDYDLIN
LKAINDCNVEDEAHVITIKLSDLAIDDKIDLFKNTNDFAIKKLIINDYTKOMLG
KLISLSTGNFNPNTIISKLEKGFQOMLNIHQCVKQKOCPENSGCFRHLDRRECKC
LLNFKQEGDKVCVENPNPTCNENGGCDADAKCTEEDSGNGKIIITCECTKPDSPYPLFD
GIFCSSENLG15FLILMLILYSFI"

38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57
Db 3859 TCCGAAGATAATGACGAATATTAGATCAAGTAACTAGCAAGAA----- 3903

58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
Db 3904 GCAATATCTGTCAAGATGGATAATATCTCTCAGGATTTGAAATGAATATGATGTTATA 3963

78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
Db 3964 TATTTAAACCTTTAGCTGGATATATAGAAGCTTAAAAACAAATTTGAAAAACATT 4023

98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
Db 4024 ATTACATTTAATTTAAATTTGACGATATCTTAAATTCAGCTCTTAAGAACGAAAAATAT 4083

118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
Db 4084 TTCTTAGATGTATTAGATCTGATTTAATGCAATTTAAACATATATCTCAATGAATAC 4143

138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
Db 4144 ATTTATCAAGATTCATTTAAATTTAATTCAGAACAAAAACACACTTTTAAAAAGT 4203

158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
Db 4204 TACAAATATATAAAGAAATCAGTAGAAAATGATATTAATTTGCACAGGAGGTAAAGT 4263

178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysValIle 197
Db 4264 TATTATGAAAAGTTTATAGCAATATATAGGATGATTAGATCAATTAATAAGTTATC 4323

198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
Db 4324 AAAGAAGAAAAAGGAG---TTCCCATCATCACCAACACACACCTCCGTCACCAAGCAAAA 4380

218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
Db 4381 ACAGACGAACAAAAAGAGGAAGTAAGTTCTCTCCATTTTAAACAAACATTGAGACCTTA 4440

238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
Db 4441 TACAAATACCTTAGTTATAAATTTGACGATTTACTTTAATTTAAAGGCAAGATTAAAC 4500

258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
Db 4501 GATTGTAATGTGAAAAAGATGAAGCACATGTTAAAAATACTAAACTTAGTGATTTAAAA 4560

278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297

Db 4561 GCAATTCATCACAAAATAGATCTTTTTAAAAACACTAACGACTTCGAAGCAATTAATAAAA 4620

QY 298 LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317

Db 4621 TTGATAAATGATGATACGAAAAAGATATCTCTGGCAAAATTTACTTAGTACAGGATTAGTT 4680

QY 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337

Db 4681 CAAAATTTTCTTAATAATAATAATCAAAATTAATTTCAAGGAAAAATTTCCAAGATATCTTA 4740

QY 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357

Db 4741 AACATTTCAACACCAATCGTAAAAAAAACAAATGTCGAGAAAAATTTCTGGATGTTTCAGA 4800

QY 358 HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys 377

Db 4801 CATTTAGTAGAAGAGAAAGATGTAATGTTTATTAATTTACAAACAAAGAGGTGATATAA 4860

QY 378 CysValGluAsnProAsnProThrCysAsnGluAsnGlnGlyCysAspAlaAspAla 397

Db 4861 TGTGTTGAAAAATCCAAATCTCTACTTGTACGAAAAATAATGTTGATGTCAGATGCC 4920

QY 398 ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys 417

Db 4921 AAATGTACCCGAAGAAGATTCAGGTAGCAACGAGAAAGAAATACATGTGAATGACTATAA 4980

QY 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer 431

Db 4981 CCGATTCCTTATCACCTTTTCGATGATGTTTTCGCACTGCC 5022

RESULT 35
AF062348 5243 bp DNA linear INV 09-FEB-2001
LOCUS Plasmodium falciparum strain HNI merozoite surface protein 1
DEFINITION precursor (msp1) gene, complete cds.
ACCESSION AF062348
VERSION AF062348.1 GI:3859842
KEYWORDS Plasmodium falciparum.
SOURCE Plasmodium falciparum.
ORGANISM Plasmodium falciparum.
REFERENCE 1 (bases 1 to 5243)
AUTHORS Jiang,G., Liu,R., Daubenberger,C.A. and Pluschke,G.
TITLE Sequence analysis of the MSP 1 gene of Plasmodium falciparum from Hainan, China
JOURNAL Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung Ping Tsa Chih 17 (5), 294-297 (1999)
REFERENCE 2 (bases 1 to 5243)
AUTHORS Jiang,G., Liu,R., Daubenberger,C.A. and Pluschke,G.
TITLE Direct Submission
JOURNAL Submitted (30-APR-1998) Molecular Immunology, Swiss Tropical Institute, Socinstrasse 57, Basel CH-4002, Switzerland
FEATURES
Location/Qualifiers
1..5243
/organism="Plasmodium falciparum"
/strain="HNI"
/db_xref="taxon:5833"
/chromosome="9"
/country="China"
<66..>5150
/gene="msp1"
<66..>5150
/gene="msp1"
/product="merozoite surface protein 1 precursor"
66..5150
/gene="msp1"
/note="binds spectrin"
/codon_start=1
/product="merozoite surface protein 1 precursor"
/protein_id="AAC72884.1"
/db_xref="GI:3859843"
/translation="MKIIFLCSFLFIINTQCVTHTESYQELYKKLEALEDAVLTGYS"


```

Alignment Scores:
Pred. No.: 5, 2e-125 Length: 5243
Score: 1915.00 Matches: 370
Percent Similarity: 95.94% Conservativeness: 10
Best Local Similarity: 93.91% Mismatches: 10
Query Match: 84.18% Indels: 6
DB: 3 Gaps: 2

US-10-057-531A-2 (1-431) x AF062348 (1-5243)

Qy 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3924 TCCGAAGATAATGACGAATATTTAGATCAAGTACTAGTGGAGAA----- 3968

Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValle 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3969 GCAATATCTGTCACAATGGGATAATATCTCTCAGAGATTGAAAAATGAATATGATGTTATA 4028

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4029 TATTTTAAACCTTTAGCTGGAGTATATAGAGACCTTTAAAAAACCAATTTGAAAAAACAAAT 4088

Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4089 ATTACATTTTAAATTTGAACGATATCTTAAATTCACGCTCTTAAGAAACGAAAAATAT 4148

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4149 TTCTTAGATGTTTAGAATCTGATTTAATGCAATTTAAACATATATATCCTCAAAATGAATAC 4208

Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4209 ATTATGGAAGATTTCATTTTAAATTTATGAAATTCAGAACAAAAAACACACTTTTAAAAAGT 4268

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyLeSer 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4269 TACAAATATATAAAGAATCAGTAGAAAATCATATTAATTTGCACAGGAGGTATTAAGT 4328

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysLysValle 197
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4329 TATTATGAAAAGGTTTTAGCGAAATATTAAGGATGATTTAGAAATCAAAATTAAGGTTATC 4388

```

/organism="Plasmodium falciparum"
 /strain="HN2"
 /db_xref="taxon:5833"
 /chromosome="9"
 /country="China"
 <66..>5180
 /gene="mspl"
 <66..>5180
 /gene="mspl"
 /product="merozoite surface protein 1 precursor"
 66..5180
 /gene="mspl"
 /note="binds spectrin"
 /codon_start=1
 /product="merozoite surface protein 1 precursor"
 /protein_id="AAC72885.1"
 /db_xref="GI:3859845"
 /translation="MKIIFLCFLPFIINTQCVTHESYQELVKKLEALEDAVLGYS
 LFOKEKRVNLEGGTGAATVSTPGSGGSGSVASGGSGSVASVSGSVASVAS
 GGSGNRVTPNSDSDAKSYADLKHRYNYLFTIKELKYPELEDTLTHMLTLCND
 IHGFKYLIDGEEINELLYKLNFYDLRAKLNDACANSYCOIPFNKIRANDELVLK
 KIVFGKPLDNIKNVGMEDYIKKNKTTIANINELIEGSKTIDONKNADNEBGGK
 KLYOAVNLFIYNKLOEAKHLISVLEKRTDILKKNENIKKLLDKINEIKNPANSG
 NPTNLLDKKLEEEHEEKIKELAKTIKFNISLFDPLDLELYLREKKNKVDYTPKS
 QDPTKSVQIPKVPNGIYVPLPLTDHNSLAADNKNSTGDLMDNPTKEIKNEKIT
 DNKERIFINIKKQIDLEKNINHTKEQNKLEDEYKSKDYEEELKFEYEMKFN
 NFDKVVVKIIFSARYTYNVEKYNKNSFNSSNVNVQKLLKALSYLEDYSRKGIS
 EKDFNHVYTKTLEADIKLKEIKSENKILEKNFKGLTHSANASLEYSIVKLQV
 QKVLILKLEKRLKELFLKNAOLKDSIHVPNIYKPNKPEPYLLVLEKVDKLEF
 IPKVKMLKKEQAVLSITOPVAASETTEDGSHSTHLSQSGETVETERTETVY
 GHVTVITLPPQPPPKVVENSIKHSNDNSOALTKTVYLKLDLDEFLEKSYIC
 HKILVNSMSDGLVLYNLTPPEKNELKSCDPLDLFIQNNIPAMYLSYDSMND
 LQHLFFELYQKEMIIYHLKKEENHKKLEEQKQITGTSSTSPGNTVNTAOSATH
 SNQOQSASNTYONGVAVSPPAVVEEDPLVLSISNDLKGIVSLNLGNKTK
 VNPATITSTMEKFEYENILKNDITVFNDDIKOFVKSNSKVITGLTOKNALNDEK
 KLKDTQLSDLYNKYKLDLRFNKKELGQDKMOIKKLTLLKEOLSKLSLNPH
 NVLQNSVFENKKEAEIAENTLNTKILKHYGLVLYNGESPLKTLSEVSIQ
 TEDYANLEKFRVLSKIDGLNHLKGLKLSFLSGSLHLLELKEVETKKNYTG
 SPSENKKNEALKSVLPEAFKVTIVTPQPDVTPSPSVRSVSGSGSTKEETQI
 PTSGSLTELQVQVQSONYDEEDSLVLPFIQSEEDNDEYLDQVTKGEAIVTMDNI
 LSGFENEYDIYPLKAGVYRSKQIKERNITFNMLNDILNSRLKRRYFDVLES
 DLMQFHISSEYIIEESFKLNSQKNTLLSKYIKESVENDIKFAOEGISYKVK
 LAKYKDLSEIKVIKEEFEPSPPTTPSPAKTDEQKESKFLPLNTFELYNNL
 VNKIDYLYNLKAKINDCNVEKDEAHVTKLSDLKAIIDKIDLFKNTDFEAKKLI
 NDKDKMLKGLSTGLQVQNPNTIISKLIEGRFQDMNISQHCYKVKQCPNSGCFR
 HLDREBECKLLYKQEGDKVCPNPTENNGGCDADAKTEEDSGSGSKKITCEC
 TKPDSPVLFQIFCSSLNFIQLLILMLILYSFI"

gene

mRNA

CDS

Db	4119	ATTACATTAATAATTTAAATTTGAACGATATCTTAAATTCACGCTCTTAAGAAACGAAATAT	4178
Qy	118	PhelLeuaspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr	137
Db	4179	TTCTTAGATGATTAGAAATCTGATTTAATGCAATTTAAACATATATCTCAATGAATAC	4238
Qy	138	IleileGluaspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer	157
Db	4239	ATTATTGAAGATTTCATTTAAATTTGAATTCAGAACAAAAACACACTTTTAAAGAT	4298
Qy	158	TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluIleSer	177
Db	4299	TACAAATATATAAAGAAATCAGTAGAAATGATATAATTTGACAGGAAGGTATAAGT	4358
Qy	178	TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysValIle	197
Db	4359	TATTATGAAAGGTTTGTAGCGAAATATAAGGATGATTTAGAAATCAATTTAAAGATTTATC	4418
Qy	198	LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys	217
Db	4419	AAAGAAGAAAAAGGAG---TTCCCATCATCACCAACCAACACCTCCGTCACCAAGCAAAA	4475
Qy	218	ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu	237
Db	4476	ACAGACCAACAAAAAGAGAAAGTAAGTTCTTCCATTTTAAACAAACATTTGAGACCTTA	4535
Qy	238	TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleasn	257
Db	4536	TACAAATTAATTAATAAATTTGACGATTACCTTAATTAATTAACGCAAAAGATTAAAC	4595
Qy	258	AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys	277
Db	4596	GATTGTAATTTGAAAAAGATGAAGACACATCTTAAATAAATAAATTTAGTGTATTAATA	4655
Qy	278	AlaIleaspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys	297
Db	4656	GCAATTCATGACAAAAATAGATCTTTTAAAAAACAACACTAACGACTTCGAAGCAATTTAAAAA	4715
Qy	298	LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal	317
Db	4716	TTGATAATGATGATACGAAAAAGATATGCTTGGCAAAATTTACTTAGTACAGGATTAGTT	4775
Qy	318	GlnAsnPheProAsnThrIleIleSerLysLeuIleGluLysPheGlnAspMetLeu	337
Db	4776	CAAAATTTTCTTAATAACAATAATCAAAATTAATTAATTAATTAATTAATTAATTAATTAAT	4835
Qy	338	AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPhearg	357
Db	4836	AACATTTTCACAAACCAATCGTAAAAAACAATGTCAGAAAAATTCGGATGTTTTCAGA	4895
Qy	358	HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys	377
Db	4896	CATTTCAGATGAACAGAGAAGATGTAATGTTTATTAAATTAACAACAAGAGGTGATAAA	4955
Qy	378	CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla	397
Db	4956	TGTGTTGAAAAATCCAAATCTCTACTGTAAACGAAAAATTAATGTTGATGTCAGATGCC	5015
Qy	398	ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluCysThrLys	417
Db	5016	AAATGTACCGAAGAAGATTTCAGGTAGCAACGGAAGAAAAATCACATGTGAATGTACTAAA	5075
Qy	418	ProAspSerTyrProLeuPheAspGlyIlePheCysSerSer	431
Db	5076	CTCGATTCTTATCCACTTTCGATGGTATTTCTTCGACTTCC	5117

RESULT 37

PFU20730

LOCUS

DEFINITION

ACCESSION

1116 bp DNA linear INV 14-FEB-1995
 Plasmodium falciparum clone B16A merozoite surface protein 1
 (MSP-1) gene, 42 kDa C-terminal region, partial cds.
 U20730

Alignment Scores:

Pred. No.: 5,23e-125 Length: 5271
 Score: 1915.00 Matches: 370
 Percent Similarity: 95.94% Conservative: 8
 Best Local Similarity: 93.91% Mismatches: 10
 Query Match: 84.18% Indels: 6
 DB: 3 Gaps: 2

US-10-057-531a-2 (1-431) x AF062349 (1-5271)

Qy	38	ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet	57
Db	3954	TCCCAAGATAATGAGTAATTTAGATCAAGTAGTACTGAGAA-----	3998
Qy	58	AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle	77
Db	3999	GCAATATCTGTCACATGATAATATCTCTCAGGATTTGAAATGAATGATGTTATA	4058
Qy	78	TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle	97
Db	4059	TATTAAACCTTTAGCTGGAGTATAGAGACCTTAAACAAACAAATTTGAAAAACATT	4118
Qy	98	PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr	117

```

VERSION      U20730.1  GI:564861
SOURCE       malarial parasite.
ORGANISM     Plasmodium falciparum
REFERENCE    1 (bases 1 to 1116)
AUTHORS      Shi.Y.-P., Alpers.M.P., Povoa.M.M., Nahlen,B.L., Oloo.A.G. and
              Lai.A.A.
TITLE        Sequence of the C-terminal region of merozoite surface protein-1
              (MSP-1) in field-derived Plasmodium falciparum
JOURNAL      Unpublished
AUTHORS      Shi.Y.-P.
TITLE        Direct Submission
JOURNAL      Submitted (07-FEB-1995) Ya-ping Shi, DPD/CDC 4770, Molecular
              Vaccine Section, Bldg 22, Rm. 8F12, Buford Highway, Chamblee, GA,
              30341 USA
FEATURES     Location/Qualifiers
             1..1116
               /organism="Plasmodium falciparum"
               /isolate="Infected human blood sample from Papua New
               Guinea"
               /db_xref="taxon:5833"
               /clone="B16A"
             gene      1..1116
               /gene="MSP-1"
             CDS       1..1116
               /gene="MSP-1"
               /codon_start=1
               /product="merozoite surface protein 1, 42 kDa C-terminal
               region"
               /protein_id="AA62221.1"
               /db_xref="GI:564862"
               /translation="AISVTMDNILSGFENEYDVIYKPLAGVYRSLLKKQIEKNITFN
               LNLNDILNSRLKRVFLDVLDMOPKHISSEYIIEDESKLLNSEOKNTLLKSYK
               YTKESVDNIEKAQEGISYVEKVLAKYDDLESIKKVIKEKEFPSSPTTPPSAKT
               DQKESKFLPELNIETRIYNNLVNKIDYLLNKAINDCNVERNEARVTKLSDL
               KAIIDKIDLFKNHDFEAIKLIINDTKMDLGLSLTGVLQNFNTIISKLEIKFQ
               NDLNTSQHQQKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNGG
               CDADAKCTEEDSGNGKKITCECTKPDSPYPLFDGIFCS"
BASE COUNT   486 a 145 c 160 g 325 t
ORIGIN
Alignment Scores:
Pred. No.:    4.05e-125      Length: 1116
Score:        1906.50      Matches: 366
Percent Similarity: 98.66%  Conservative: 2
Best Local Similarity: 98.12% Mismatches: 4
Query Match:   83.80%      Indels: 1
DB:            3           Gaps: 1

US-10-057-531A-2 (1-431) x PFU20730 (1-1116)

QY  58  AlaileSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
    |||
Db  1  GCAATATCTGTCACATGGATATATCCCTCAGGATTTGCAAAATGAATATGATGTTATA 60
    |||

QY  78  TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
    |||
Db  61  TATTATAAACCTTTAGCTGGAGTATATAGAAAGCTTAAAAAACAATAATGAAAAACATT 120
    |||

QY  98  PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
    |||
Db  121  TTTTACATTTAAATTTGAACGATATCTTAAATTCAGCTCTTAAAGAACGAAAAATAT 180
    |||

QY  118  PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
    |||
Db  181  TTTCTTAGATCTATTAGAAATCTGATTAATCCATTTAAACATATATATCCCTCAATGAATAC 240
    |||

QY  138  IleileGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
    |||
Db  241  ATTATTGAAGATTCAATTAATTTATTGAATTCAGAACAAAAACACACACTTTTAAAAAGT 300
    |||

```

```

QY  158  TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
    |||
Db  301  TACAATAATATAAAGAATCAGTAGAAAATCATATTAAATTTGCACAGGAGGTATAAGT 360
    |||

QY  178  TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysValIle 197
    |||
Db  361  TATTATCAAAAGGTTTGTAGCGAAATATAAGGATGATTAGAACATTAATAAAGATTATC 420
    |||

QY  198  LysGluGluLysGluLysPheProSerSerProThrThrProProSerProAlaLys 217
    |||
Db  421  AAAGAAGAAAAAGGAG--TTTCCCATCATCACCAACAACACTCCGTCACCAGCAAAA 477
    |||

QY  218  ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
    |||
Db  478  ACAGACGAACAAAAAGAGAAAGTAAAGTTCTCTTCCATTTTAAACAAATTTGAGACCTTA 537
    |||

QY  238  TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
    |||
Db  538  TACATAAATCTAGTTAATAAATTTGAGGATTACTTAATTAACTTAAAGGCAAGATTAAAC 597
    |||

QY  258  AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
    |||
Db  598  GATTGTAATGTTGAAAAAATGAAGCACATGTTAAAAATACTAAACTTAGTGATTTAAAA 657
    |||

QY  278  AlaileAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
    |||
Db  658  GCAATTTGATGACAAAATAGATCTTTTTAAAAACCATTAACGACTTCGAGCAATTTAAAA 717
    |||

QY  298  LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal 317
    |||
Db  718  TTGATAAATGATGATGAAAAAAGATATGCTTGGCAAAATTAATTAATTAACGAGGATTAGT 777
    |||

QY  318  GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
    |||
Db  778  CAAATTTTCTTAATCAATAATATCAAAATTAATTAATTAACGAGAAATTTCCAAGATATCTTA 837
    |||

QY  338  AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
    |||
Db  838  AACATTTTCAACACCAATCGTAAAAAACAATGTCCAGAAAAATTTCTGGATGTTTTCAGA 897
    |||

QY  358  HisLeuAspGluArgGluGluCysLysCysLeuLeuAsnTyrLysGlnGluLysAspLys 377
    |||
Db  898  CATTTAGATGAAAGAGAAATGTAATGTTTATTAATTAACAAACGAGGAGGTGATATA 957
    |||

QY  378  CysValGluAsnProAsnProThrCysAsnGluAsnGlnGlyCysAspAlaAspAla 397
    |||
Db  958  TGTGTTGAAAAATCCAAATCTACTTGTAAACGAAAAATTAATGTTGGATGTCAGATGCC 1017
    |||

QY  398  ThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys 417
    |||
Db  1018  AAATGTACCGAAGAAGATTTCAGGTAGCAACGAAAGAAAAATTCACATGTGAATGTACTATA 1077
    |||

QY  418  ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
    |||
Db  1078  CCTGATTTCTTATCCACTTTTCGATGGTATATTTCTGTCAGT 1116
    |||

RESULT 38
PFU20731
LOCUS
DEFINITION
  PFU20731 1116 bp DNA linear INV 14-FEB-1995
  Plasmodium falciparum clone B16C merozoite surface protein 1
  (MSP-1) gene, 42 kDa C-terminal region, partial cds.
ACCESSION
  U20731
VERSION
  U20731.1
SOURCE
  U20731.1 GI:564863
  malarial parasite.
  ORGANISM
  Plasmodium falciparum
  Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
  REFERENCE
  1 (bases 1 to 1116)
  AUTHORS
  Shi.Y.-P., Alpers.M.P., Povoa.M.M., Nahlen,B.L., Oloo.A.G. and
  Lai.A.A.
  TITLE
  Sequence of the C-terminal region of merozoite surface protein-1
  (MSP-1) in field-derived Plasmodium falciparum
  JOURNAL
  Unpublished

```

```

REFERENCE 2 (bases 1 to 1116)
AUTHORS Shi,Y.-P.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-1995) Ya-Ping Shi, DPD/CDC 4770, Molecular
Vaccine Section, Bldg 22, Rm. 8F12, Buford Highway, Chamblee, GA,
30341 USA
FEATURES
source Location/Qualifiers
1..1116
/organism="Plasmodium falciparum"
/isolate="infected human blood sample from Papua New
Guinea"
/db_xref="taxon:5833"
/clone="B16C"
1..1116
/gene="MSP-1"
<1..>1116
/gene="MSP-1"
/codon_start=1
/product="merozoite surface protein 1, 42 kDa C-terminal
region"
/protein_id="AAA62222.1"
/db_xref="GI:564864"
/translation="AISVTMDHILSGFENEVDVYLKPLAGVYRSLKKOIEKNIFTFN
LNLDILNSRLKRYKYLVDLESLMOFKHLSSEYIIESEFKLLNSEKNTLLKSKY
YIKESVENDIKPAGEGISYEKVLAKYKDDLESKKVKEKEFPSPPTPPSPAKT
DEQKESKFLPFLNIETLYNLNLIKNDKNCVENKNEAHKVKITKLSDL
KAIDKIDLFKNHNDFFAIKKLIINDTKDKMLGLLSTGLVQNPNTIISKLIIEGRFQ
DMLNISQHCVKQCPENSGCFRHLDERECKILLNYKQEGDKCVENPNPTCENNGG
CDADAKTEEDSGNSGKKITCECTKPDSPFLFDGIFCS"
BASE COUNT 485 a 146 c 160 g 325 t
ORIGIN
Alignment Scores:
Pred. No.: 9.07e-125 Length: 1116
Score: 1901.50 Matches: 365
Percent Similarity: 98.66% Conservative: 3
Best Local Similarity: 97.86% Mismatches: 4
Query Match: 83.58% Indels: 1
DB: 3 Gaps: 1
US-10-057-531a-2 (1-431) x PFU20731 (1-1116)
Qy 58 AlalieSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
|||||
Db 1 GCATATCTGTCACAAATGATCATATCTCTCAGGATTGTGAAATGAATATGATGTTATA 60
Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
|||||
Db 61 TATTTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAAACAATTTGAAAAACATT 120
Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
|||||
Db 121 TTTACATTTAATTTAAATTTGAACGATATCTTAAATTCACGCTCTTAAGAACGAAATAT 180
Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
|||||
Db 181 TTTCTAGATGTTAGATCTGATTTAATGCAATTTAAACATATATCTCAATGAATAC 240
Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLysSer 157
|||||
Db 241 ATTATTGAAGATTCATTTAAATTTAGATTCAGACAAAAAACAACACTTTTAAAAAGT 300
Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluLysSer 177
|||||
Db 301 TACAAATATATAAAGAATCAGTACAAAATGATATTAAATTTGCACAGGAGGTATAAGT 360
Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspLeuGluSerIleLysValIle 197
|||||
Db 361 TATTATGAAAGGTTTACGCAAAATATAGGATGATTTAGATCAATTTAAAAAGTTATC 420
Qy 198 LysGluGluLysGluLysPheProSerProThrProThrProProSerProAlaLys 217
|||||
Db 421 AAAGAAGAAAGGAG---TTCCCATCATCACCAACCAACACCTCCCTCACCAGCAAAA 477

```

```

Qy 218 ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu 237
|||||
Db 478 ACAGACGACAAACAAAGAGAAAGTAAGTTCCTTCATTTTAAACAAACATTGAGACCTTA 537
Qy 238 TyrAsnAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn 257
|||||
Db 538 TACAATAACTTTAGTTAATAAATTCACGATTACTTTAATTAACCTTAAGGCAAGATTAAAC 597
Qy 258 AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys 277
|||||
Db 598 GATTGTAATGTTGAAAAAATGAACACATGTTAAAAATACTAACTTAGTGATTATAAA 657
Qy 278 AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaIleLysLys 297
|||||
Db 658 GCAATTGATGACAAATAGATCTTTTAAACCAATACGACTTCGAGCAATTAAGAAA 717
Qy 298 LeuIleAsnAspAspThrLysLysAspMetLeuGluLysLeuLeuSerThrGlyLeuVal 317
|||||
Db 718 TTGATAAATGATGATACGAAAAAAGATATGCTTGGCAAAATTACTTAGTACAGGATTAGT 777
Qy 318 GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu 337
|||||
Db 778 CAAAATTTCTTAATACAAATATATCAAAATTAATTGAAGGAAAAATTCGAAGATATGTTA 837
Qy 338 AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg 357
|||||
Db 838 AACATTTCCAAACACCAATGCGTAAAAAACAATGTCAGAAAAATCTCGGATGTTTCAGA 897
Qy 358 HisLeuAspGluAtrGluGluCysLysCysLeuLeuAsnTyrLysGlnGluLysAspLys 377
|||||
Db 898 CATTTAGATGAAGAAGAAATGTAATGTTTATTAATTAACAAAGAGGTGATATAA 957
Qy 378 CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla 397
|||||
Db 958 TGTGTTGAAATCCAAATCCTACTTGTAAACGAAATAATGTTGGATGTGATGCAGATGCC 1017
Qy 398 ThrCysThrGluGluAspSerGlySerSerArgLysLysIleThrCysGluCysThrLys 417
|||||
Db 1018 AAATGTACCCGAAGAAGATTTCAGGTAGCAACGGAAGAAAAATCACATGTGAATGTAATAA 1077
Qy 418 ProAspSerTyrProLeuPheAspGlyIlePheCysSer 430
|||||
Db 1078 CCTGATCTTATCCACTTTTCAGTGGTATTTCTGCACT 1116
RESULT 39
AF040569 1116 bp DNA linear INV 13-JUN-1998
LOCUS Plasmodium falciparum strain Kenya-3 merozoite surface protein-1
DEFINITION (MSP-1) gene, partial cds.
ACCESSION AF040569
VERSION AF040569.1 GI:2789669
KEYWORDS
SOURCE Plasmodium falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 1116)
AUTHORS Qari,S.H., Shi,Y.P., Goldman,I.F., Nahlen,B.L., Tibayrenc,M. and Lal,A.A.
TITLE Predicted and observed alleles of Plasmodium falciparum merozoite
surface protein-1 (MSP-1), a potential malaria vaccine antigen
JOURNAL Mol. Biochem. Parasitol. 92 (2), 241-252 (1998)
MEDLINE 98319411
PUBMED 9657329
REFERENCE 2 (bases 1 to 1116)
AUTHORS Qari,S.H., Shi,Y.P., Goldman,I., Nahlen,B., Tibayrenc,M. and Lal,A.A.
TITLE Direct Submission
JOURNAL Submitted (31-DEC-1997) DASTLR, Centers for Disease Control &
Prevention (CDC), 1600 Clifton Road, Atlanta, GA 30333, USA
FEATURES
Location/Qualifiers
source 1..1116
/organism="Plasmodium falciparum"

```

```

/strain="Kenya-3"
/isolate="10"
/db_xref="taxon:5833"
/chromosome="9"
<1..>1116
/gene="MSP-1"
/allele="Kenya-3"
<1..>1116
/gene="MSP-1"
/feature="42 kDa C-terminal region"
/codon_start=1
/product="merozoite surface protein-1"
/protein_id="AAC39099.1"
/db_xref="GI:2789670"
/translation="AISVTMDNIIISGFENEDYIVLKLPLAGYRSLKQIKENIPTFN
LNLINDLRLKRYFLVDLQKHLSSNEYIIEDESKLLNSKQKNTLLKSYK
YIKESVENDIKPAQEGISYEKVLAKYKDDLESKKVKEKEPSPPTPPSPAKT
DEQKESKFLPFLNIEYLNKIDHYLNKAKINDCNVEKDEAHVITKLSDL
KALDDKIDLFKNHNDFAIKKLLINDDKKMLLSTGLVONFPNTIISKLEKFKQ
DMLNISQHCYKQCPENSGCFRHLDERECKLLNLYKQEGDKCVENPNPTCNENNGG
CDAKACTEEDSGSKKITCECTKPDSPYFFDGFCS"
BASE COUNT 485 a 145 c 160 g 326 t
ORIGIN

```

Alignment Scores:
 Pred. No.: 1.07e-124 Length: 1116
 Score: 1900.50 Matches: 365
 Percent Similarity: 98.12% Conservative: 1
 Best Local Similarity: 97.86% Mismatches: 6
 Query Match: 83.54% Indels: 1
 DB: 3 Gaps: 1

US-10-057-531A-2 (1-431) x AF040569 (1-1116)

Qy	58	AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle	77
Db	1	GCAATATCTGTCAATGGATATATCTCTCAGGATTTTGAATAATGAATATGATTTATA	60
Qy	78	TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysGlyGlnIleGluLysAsnIle	97
Db	61	TATTTAAACCTTTAGCTGGATATATAGAGCTTAAAAAACAATTCGAAATTAATGATTTATA	120
Qy	98	PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr	117
Db	121	TTTACATTTAAATTTAAATTTGAACGATATCTTAAATTCACGCTTAAAGAAACGAAATAT	180
Qy	118	PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr	137
Db	181	TTCTTAGATGTTATAGATCTGATTTAATGCAATTTAAACATATATCTTCAATGAATAC	240
Qy	138	IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer	157
Db	241	ATTATTGAAGATTCATTTAAATTTAATTTGAATTCAGAACAAAAAACACACTTTTAAAGCT	300
Qy	158	TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer	177
Db	301	TACAAATATATAAAGAATCAGTAGAAATGATATTAATTTTGCACAGGAAGGTATAAGT	360
Qy	178	TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysValIle	197
Db	361	TATATGAAGAGTTTACGGAATATAAGGATGATTAGATCAATTAATAAGGTATATC	420
Qy	198	LysGluGluLysGluLysPheProSerSerProThrThrProProSerProAlaLys	217
Db	421	AAAGAAGAAAGAG--TTCCCATCATCACCAACCAACACCTCCGTCACCAAGCAAA	477
Qy	218	ThrAspGluGlnLysLysGluSerLysPheLeuProPheLeuThrAsnIleGluThrLeu	237
Db	478	ACAGACGAACAAAGAAAGAAAGTTCCTTCCATTTTAAACAACATTTGAGACCTTA	537
Qy	238	TyrAspAsnLeuValAsnLysIleAspAspTyrLeuIleAsnLeuLysAlaLysIleAsn	257
Db	538	TACAATAACTTATGTTAATAAATTTGACCATTTACTTAACTTAAAGGCAAGATTAAC	597

Qy	258	AspCysAsnValGluLysAspGluAlaHisValLysIleThrLysLeuSerAspLeuLys	277
Db	598	GATTGTAAATGTTCAAAACAGATGAAGCACATGTTAAATAACTAACTTACTGATTTAAA	657
Qy	278	AlaIleAspAspLysIleAspLeuPheLysAsnProTyrAspPheGluAlaLysLys	297
Db	658	GCAATTTGATGACAAAATAGATCTTTTAAAAACCAATACGACTTCGAAGCAATTAATA	717
Qy	298	LeuIleAsnAspAspThrLysLysAspMetLeuGlyLysLeuLeuSerThrGlyLeuVal	317
Db	718	TTGATAAATGATGATACGAAAAGATATGCTTGGCAAAATTTACTTAGTACAGGATAGTT	777
Qy	318	GlnAsnPheProAsnThrIleIleSerLysLeuIleGluGlyLysPheGlnAspMetLeu	337
Db	778	CAAAATTTTCTTAATAACAATAATATCAAAATTAATTAAGGAAGAAATTCGAAGATATGTTA	837
Qy	338	AsnIleSerGlnHisGlnCysValLysLysGlnCysProGluAsnSerGlyCysPheArg	357
Db	838	AAATTTTCAACACCAATCGCTTAAAAAACCAATGTCCAGAAATTTCTGGATGTTTTCAGA	897
Qy	358	HisLeuAspGluArgGluGlyCysLysCysLeuLeuAsnTyrLysGlnGluGlyAspLys	377
Db	898	CATTTAGATGAAAGAGAAGAAATGTAATGTTTAAATTTACAAACAAAGAGGTGATAA	957
Qy	378	CysValGluAsnProAsnProThrCysAsnGluAsnAsnGlyGlyCysAspAlaAspAla	397
Db	958	TGTGTTGAAATCCAAATCTACTTCTAACGAAATAATGGTGGATGATGATGATGCC	1017
Qy	398	ThrCysThrGluGluAspSerGlySerArgLysLysIleThrCysGluGlyThrLys	417
Db	1018	AAATGTACCGAAGAGATTCAGGTACGAGGAAAGAAATACATGCTGAATGATCTATAA	1077
Qy	418	ProAspSerTyrProLeuPheAspGlyIlePheCysSer	430
Db	1078	CCTGATTCATTCATTTTCGATGATTTTCTGAGT 1116	
RESULT 40			
PFAMSP807			
LOCUS			
DEFINITION			
P. falciparum DNA for the precursor to the major merozoite surface			
proteins, C-terminal.			
ACCESSION			
D13344			
VERSION			
D13344.1 GI:391786			
KEYWORDS			
EGF-like domains; major merozoite surface protein precursor.			
SOURCE			
Plasmodium falciparum (isolate 807 and 836) merozoite, DNA.			
ORGANISM			
Plasmodium falciparum			
REFERENCE			
1 (bases 1 to 1726)			
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
AUTHORS			
Jongwutives, S., Tanabe, K. and Kanbara, H.			
TITLE			
Sequence conservation in the C-terminal part of the precursor to			
the major merozoite surface proteins (MSP1) of Plasmodium			
falciparum from field isolates			
JOURNAL			
Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)			
MEDLINE			
93295445			
REFERENCE			
2 (bases 1 to 1726)			
Jongwutives, S.			
AUTHORS			
Direct Submission			
TITLE			
Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical			
Medicine, Nagasaki Univ., Dept. of Protozoology, 12-4			
Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747),			
Fax:0958-47-6607)			
COMMENT			
Submitted (05-OCT-1992) to DDBJ by:			
Somchai Jongwutives			
Department of Protozoology			
Institute of Tropical Medicine			
Nagasaki University			
12-4 Sakamoto-machi			
Nagasaki 852			
Japan			
Phone: 0958-47-2111x3747			
Fax: 0958-47-6607.			
Location/Qualifiers			
FEATURES			

```
source      1. .1726
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="9"
/dev_stage="merozoite"
gene        1. .1712
/gene="MSPI"
<1. .1712
CDS
/gene="MSPI"
/codon_start=3
/product="major merozoite surface protein precursor"
/protein_id="BA02605.1"
/db_xref="GI:391787"
/translation="PLKTLSEVSIQTDNYANLEKFRVLSKIDGKLNLDHLGKKLS
FLSSGHLHILTELKEVINKNYTGNSPNNKKNVNEALKSYENFLPEAKVTVTTPQ
PDVTPSPVSRVSGSGSTKEETIPTSGSLLELOQVQSONYDEEDSKVLPLPG
ESPDNEYLDQVVTGSAISVTMDNIIISGRENEDVILKPLAGVYBSLKQKLEKNIIT
FNLNLDLNSRLKRRKYFLDVLSDLMOFKHISSENEYIEDSFKLNSEOKNLTLS
YKIKESVENDIKFADEGISYEKVLAKYKDDLESKKVKEKEPEPPPTPPSPA
KTDQKESKFLPFLNIETLYNNLVNKTDGDIYNLKKAKINDCNVKEDEAHVKITLS
DLKAIIDKIDLFKNTDDEFAIKKLIINDTKKMLGKLLSTGLVQNPENIISKLIEGK
FQMLNISQHCQVKOCPONSGCFRHLDERECKCLNLYKQEGDKGVENPNPTCNEN
GGCDADAKCTEEDSGSGKKTCECTKPDSPYPPFDGIFCSSSNFLGIFLLILLY
SFI"
```

BASE COUNT 739 a 233 c 246 g 508 t

ORIGIN

Alignment Scores:

Pred. No.:	1-82e-124	Length:	1726
Score:	1900.00	Matches:	367
Percent Similarity:	95.43%	Conservative:	9
Best local Similarity:	93.15%	Mismatches:	12
Query Match:	83.52%	Indels:	6
DB:	3	Gaps:	2

US-10-057-531A-2 (1-431) x PFAMSP807 (1-1726)

```
Qy 38 ThrAspAspAspLysAlaMetAlaAspIleGlySerIleGluGlyArgGlyThrMet 57
      ::::::::::::::: :::: ::::
Db 486 TCCGAAGATAATGACGAATATTAGATCAAGTAGTAACTGGAGAA----- 530

Qy 58 AlaIleSerValThrMetAspAsnIleLeuSerGlyPheGluAsnGluTyrAspValIle 77
      ::::::::::::::: :::: ::::
Db 531 GCRAATCTGTGCAATGGGATATATCTCTCAGGATTTGAAATGAATATGATGTTATA 590

Qy 78 TyrLeuLysProLeuAlaGlyValTyrArgSerLeuLysLysGlnIleGluLysAsnIle 97
      ::::::::::::::: :::: ::::
Db 591 TATTTAAACCTTTAGCTGGAGTATATAGAGCTTAAAAACAAATTTGAAAAACATT 650

Qy 98 PheThrPheAsnLeuAsnLeuAsnAspIleLeuAsnSerArgLeuLysLysArgLysTyr 117
      ::::::::::::::: :::: ::::
Db 651 ATTACATTTAATTTAAATTTGAACGATATCTTAAATTCAGCTCTTAAGAAACGAAATAT 710

Qy 118 PheLeuAspValLeuGluSerAspLeuMetGlnPheLysHisIleSerSerAsnGluTyr 137
      ::::::::::::::: :::: ::::
Db 711 TTCTTAGATGTATTAGAACTGATTTTAATGCAATTTAAACATATATCTCCTCAATGAATAC 770

Qy 138 IleIleGluAspSerPheLysLeuLeuAsnSerGluGlnLysAsnThrLeuLeuLysSer 157
      ::::::::::::::: :::: ::::
Db 771 ATTATTCAAGATTTCATTTAAATTTAATTTCAATTCAGACAAAAAACACACATTTTAAAAAGT 830

Qy 158 TyrLysTyrIleLysGluSerValGluAsnAspIleLysPheAlaGlnGluGlyIleSer 177
      ::::::::::::::: :::: ::::
Db 831 TACAAATATATAAAGAAATCAGTAGAAAAATGATATTAAATTTGCACAGGAGGTATAAGT 890

Qy 178 TyrTyrGluLysValLeuAlaLysTyrLysAspAspLeuGluSerIleLysValIle 197
      ::::::::::::::: :::: ::::
Db 891 TATTATGAAAAGGTTTATAGCAATATATAGCATGATTAGATCAATTTAAAAAGTTATC 950

Qy 198 LysGluGluLysGluLysPheProSerSerProProThrThrProProSerProAlaLys 217
      ::::::::::::::: :::: ::::
Db 951 AAGAAGAAAAAGGAG---TTCCCATCATCATCACCACCAACACACCTCGCTCACCAGCAAAA 1007
```

Search completed: May 19, 2003, 15:54:17
Job time : 2641 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: May 12, 2003, 10:21:03 ; Search time 39 Seconds
(without alignments)
1472.590 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 431

Sequence: 1 MHHHHHHHSLVPRGSMKE.....TCECTKPDVPLFDGIFCSS 431

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A_Geneseq_101002.*

```
1: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	375	87.0	375	22	AA883926
2	60	13.9	116	18	AAW36103
3	60	13.9	116	18	AAW22592
4	60	13.9	127	18	AAW22593
5	60	13.9	127	18	AAW36102
6	51	11.8	54	14	AA41357
7	50	11.6	76	21	AA26110
8	50	11.6	167	22	AA85253
9	50	11.6	167	22	AA85254
10	50	11.6	289	22	AAE13033

11	50	11.6	289	23	AA47732	HOP38 #5. Helicob
12	50	11.6	402	23	ABG67266	Polythoa 2 fluores
13	50	11.6	407	23	ABG67267	Polythoa 2 fluores
14	50	11.6	407	23	ABG67270	Polythoa 2 fluores
15	50	11.6	467	21	AA785150	Mouse secreted pro
16	50	11.6	467	21	AA79067	Murine secreted pr
17	50	11.6	551	22	AA85251	Plant thioredoxin-
18	50	11.6	551	22	AA85252	Plant thioredoxin-
19	50	11.6	692	21	AA811532	SEN virus protein
20	50	11.6	824	23	AA48247	Thioredoxin/deoxyx
21	48	11.1	49	14	AA41355	MSPIEGF1B EGF1-lik
22	46	10.7	96	22	AA37608	Merozoite surface
23	46	10.7	108	22	AA37609	Merozoite surface
24	46	10.7	138	21	AA811534	SEN virus protein
25	46	10.7	284	23	AAE20460	Recombinant (SPKR)
26	46	10.7	355	20	AA90372	Merozoite surface
27	46	10.7	355	20	AA90372	Merozoite surface
28	46	10.7	361	20	AA90373	Merozoite surface
29	46	10.7	361	20	AA90373	Merozoite surface
30	46	10.7	376	20	AA90374	Merozoite surface
31	46	10.7	376	20	AA90374	Merozoite surface
32	46	10.7	408	20	AAW84315	Modified merozoite
33	46	10.7	408	20	AAW84315	Modified merozoite
34	45	10.4	1639	19	AAW54145	TrxA-rabbit tissue
35	45	10.4	59	23	AB80661	P. falciparum synt
36	34	7.9	49	14	AA811533	N-His botulinum to
37	32	7.4	55	22	AA811534	SEN virus protein
38	27	6.3	382	20	AA93952	Human mamaglobin
39	27	6.3	382	20	AA93952	Human mamaglobin
40	23	5.3	54	14	AA41356	Gaussia luciferase
41	23	5.3	1654	6	AA50777	Gaussia species CB
42	21	4.9	299	21	AA932143	MSPIEGF2A EGF2-lik
43	21	4.9	302	21	AA808323	Sequence of the p1
44	21	4.9	302	21	AA808323	Alfa transcript
45	18	4.2	328	22	AAE05749	Amino acid sequenc
46	18	4.2	328	22	AAE05749	Chimeric S peptide
47	17	3.9	350	21	AA70287	Plasmodium falci
48	17	3.9	350	21	AA70287	Plasmodium falci
49	17	3.9	350	21	AA70287	Recombinant vaccin
50	16	3.7	31	19	AAW1706	Plasmodium falci
51	15	3.5	15	18	AAW08161	Amino acid sequenc
52	15	3.5	15	18	AAW08161	Plasmodium falci
53	15	3.5	15	19	AAW79592	S-peptide with (g1
54	15	3.5	15	19	AAW79592	Purification tag o
55	15	3.5	15	21	AAW84204	BS15-peptide. Syn
56	15	3.5	15	21	AAW84204	Peptide comprising
57	15	3.5	15	21	AAW84204	S-tag sequence. U
58	15	3.5	15	21	AAW84204	Bovine pancreatic
59	15	3.5	15	21	AAW84204	Rec2 protein leade
60	15	3.5	15	21	AAW84204	S-peptide, a known
61	15	3.5	15	21	AAW84204	Leader sequence fo
62	15	3.5	15	21	AAW84204	Ribonuclease S-pep
63	15	3.5	15	21	AAW84204	Ribonuclease S-pep
64	15	3.5	15	21	AAW84204	N-terminal His-tag
65	15	3.5	15	21	AAW84204	Human and murine O
66	15	3.5	15	21	AAW84204	Recombinant lysoph
67	15	3.5	15	21	AAW84204	Integrin cytoplasm
68	15	3.5	15	21	AAW84204	N-terminal sequenc
69	15	3.5	15	21	AAW84204	His tag peptide SE
70	15	3.5	15	21	AAW84204	Hexa-His/thrombin
71	15	3.5	15	21	AAW84204	Leader sequence fo
72	15	3.5	15	21	AAW84204	Fusion peptide lea
73	15	3.5	15	21	AAW84204	Fusion peptide lea
74	15	3.5	15	21	AAW84204	OB gene-related pe
75	15	3.5	15	21	AAW84204	N-terminal extensi
76	15	3.5	15	21	AAW84204	Human and murine O
77	15	3.5	15	21	AAW84204	Immunotoxin fusio
78	15	3.5	15	21	AAW84204	Synthetic 65k:glut
79	15	3.5	15	21	AAW84204	Peptide expressed
80	15	3.5	15	21	AAW84204	His-tag leader seq
81	15	3.5	15	21	AAW84204	Toscana virus reco
82	15	3.5	15	21	AAW84204	Toxoplasma gondii
83	15	3.5	15	21	AAW84204	PEP-15b expression
84	15	3.5	15	21	AAW84204	Murine OB carboxy

84	15	3.5	43	21	AAB28471	Murine OB peptide	157	15	3.5	371	20	AAW73369	Epitope tagged TBP
85	15	3.5	43	21	AAW77730	Murine OB protein	158	15	3.5	386	23	AAE22205	His6-NLS-Cre-MTS f
86	15	3.5	110	19	AAW78908	Rat CART1 fusion p	159	15	3.5	398	19	AAW71071	Multiple scleroses
87	15	3.5	110	20	AAW5187	pET15b-rCART1 fusi	160	15	3.5	398	20	AAW99557	Protein encoded by
88	15	3.5	116	21	AAW52367	Human tyrosine kin	161	15	3.5	398	23	ABW84070	Human proBMP-2. H
89	15	3.5	121	22	AAW31608	Amino acid sequenc	162	15	3.5	409	21	AAW07702	Protein encoded by
90	15	3.5	124	16	AAW49151	Wild-type bovine p	163	15	3.5	409	22	AAW31545	Amino acid sequenc
91	15	3.5	124	18	AAW35108	Bovine pancreatic	164	15	3.5	431	22	AAE02201	Streptococcus pneu
92	15	3.5	124	20	AAW33327	Bovine pancreatic	165	15	3.5	431	22	AAW61580	FabF protein sequen
93	15	3.5	124	20	AAW81768	Bovine RNAS peptid	166	15	3.5	485	22	AAU00222	Lfn-Bcl-XL apoptos
94	15	3.5	124	21	AAW69974	RNAS protein. Bos	167	15	3.5	588	21	AAW01279	Histidine tagged C
95	15	3.5	133	20	AAW33443	Parpox virus rppv	168	15	3.5	570	21	AAW01276	Histidine tagged A
96	15	3.5	144	19	AAW47195	Herpes simplex vir	169	15	3.5	656	23	AAW75383	Immunotoxin fusion
97	15	3.5	144	21	AAW52368	Human tyrosine kin	170	15	3.5	657	23	AAW75389	Diphtheria toxin/U
98	15	3.5	149	23	AAU10045	Tm 2.2 clone/His t	171	15	3.5	662	21	AAW01273	Neisseria meningit
99	15	3.5	149	23	AAU10047	Tm 2.3 clone/His t	172	15	3.5	665	20	AAW24940	HCV NS4A-NS3 compl
100	15	3.5	149	23	AAU10049	Tm 13.17 clone/His	173	15	3.5	665	20	AAW24941	HCV NS4A-NS3 compl
101	15	3.5	149	23	AAU10051	Tm 3.4 clone/His t	174	15	3.5	665	20	AAW24942	HCV NS4A-NS3 compl
102	15	3.5	149	23	AAU10053	Tm 3.9 clone/His t	175	15	3.5	665	20	AAW24943	HCV NS4A-NS3 compl
103	15	3.5	149	23	AAU10055	Tm 7.5 clone/His t	176	15	3.5	665	20	AAW24944	HCV NS4A-NS3 compl
104	15	3.5	150	21	AAW52369	Human tyrosine kin	177	15	3.5	665	20	AAW24945	HCV NS4A-NS3 compl
105	15	3.5	150	23	AAE20141	Human tyrosine kin	178	15	3.5	665	20	AAW24946	HCV NS4A-NS3 compl
106	15	3.5	152	23	AAW50302	Factor XIIa subst	179	15	3.5	665	20	AAW24947	HCV NS4A-NS3 compl
107	15	3.5	153	22	AAW67676	Amino acid sequenc	180	15	3.5	669	23	ABW08420	Catalytic trans-s
108	15	3.5	156	23	AAW50303	Nerve growth facto	181	15	3.5	671	20	AAW24948	HCV NS4A-NS3 compl
109	15	3.5	157	21	AAW01596	Nerve growth facto	182	15	3.5	671	20	AAW24949	HCV NS4A-NS3 compl
110	15	3.5	157	22	AAW67677	Amino acid sequenc	183	15	3.5	711	19	AAW57321	Topoisomerase I ex
111	15	3.5	157	23	AAW85725	Synthetic nerve gr	184	15	3.5	948	22	AAW31611	Amino acid sequenc
112	15	3.5	160	22	AAW02200	Streptococcus pneu	185	14	3.2	18	20	AAW73374	Epitope tagged TBP
113	15	3.5	161	23	ABW81281	Human FGF-9 mutant	186	13	3.0	15	21	AAW06025	Bovine truncated S
114	15	3.5	162	23	AAW50300	Factor XIIa subst	187	13	3.0	18	10	AAW91370	Apomigrin peptide
115	15	3.5	163	22	AAW67678	Amino acid sequenc	188	12	2.8	12	20	AAW94422	MBP/BMP fusion con
116	15	3.5	166	17	AAW00525	Human obesity prot	189	12	2.8	18	13	AAW25208	Bovine seminal RNA
117	15	3.5	166	17	AAW00529	Murine obesity prot	190	12	2.8	18	21	AAW33253	hnrNP U protein 69
118	15	3.5	166	17	AAW00533	Murine obesity prot	191	12	2.8	19	21	AAW70288	EGF-like domain of
119	15	3.5	166	17	AAW00524	Murine obesity prot	192	12	2.8	20	22	AAW70818	Extension peptide
120	15	3.5	166	22	AAW02196	Staphylococcus aur	193	12	2.8	21	20	AAW18411	Polyptide which
121	15	3.5	166	23	AAW50301	Nerve growth facto	194	12	2.8	24	22	AAW70817	Sequence #2 for pa
122	15	3.5	167	17	AAW92730	Murine obesity prot	195	12	2.8	24	22	AAW33326	Aminoacyl tRNA syn
123	15	3.5	167	17	AAW00521	Human obesity prot	196	12	2.8	134	20	ABW76190	Modified polyHis-e
124	15	3.5	167	17	AAW92726	Murine obesity prot	197	11	2.6	11	22	AAW83927	Recombinant squir
125	15	3.5	167	22	AAW00517	Human obesity prot	198	10	2.3	19	16	AAW77413	Recombinant squir
126	15	3.5	167	22	AAW67679	Amino acid sequenc	199	10	2.3	19	16	AAW60976	Recombinant squir
127	15	3.5	173	23	AAU10044	Tm 2.2 clone-His t	200	9	2.1	14	23	AAW77404	Amino acid sequenc
128	15	3.5	173	23	AAU10046	Tm 3.4 clone/His t	201	9	2.1	16	23	AAW98499	His-tag peptide SE
129	15	3.5	173	23	AAU10050	Tm 3.9 clone/His t	202	9	2.1	20	23	ABW76190	Polyhistidine-ente
130	15	3.5	173	23	AAU10052	Tm 7.5 clone/His t	203	9	2.1	21	18	AAW23647	Polyhistidine-ente
131	15	3.5	173	23	AAU10054	Tm 13.17 clone/His	204	9	2.1	21	18	AAW23650	Met Gly His10 pept
132	15	3.5	174	23	AAW76232	Salmonella Sef14 p	205	9	2.1	21	18	AAW23653	12 amino acid ente
133	15	3.5	180	19	AAW47081	Hexa-His-tagged hu	206	9	2.1	22	22	AAW67674	Enterokinase cleav
134	15	3.5	181	23	AAU10048	Hepatitis C virus	207	9	2.1	22	23	AAW49518	Peptide for expres
135	15	3.5	189	22	AAE09217	HCV NS4A-NS3 compl	208	9	2.1	24	18	AAW19779	ARH-III angiogenin
136	15	3.5	206	21	AAW44729	HCV NS4A-NS3 compl	209	9	2.1	24	20	AAW06647	HFRT peptide. Syn
137	15	3.5	213	20	AAV17888	HCV NS4A-NS3 compl	210	9	2.1	24	20	AAW27015	HFRT peptide. Syn
138	15	3.5	215	20	AAV17887	HCV NS4A-NS3 compl	211	9	2.1	24	20	AAW09370	Recombinat antic
139	15	3.5	215	20	AAV17890	HCV NS4A-NS3 compl	212	9	2.1	24	21	AAW81959	Recombinat neutro
140	15	3.5	216	20	AAV17879	HCV NS4A-NS3 compl	213	9	2.1	24	22	AAW47057	S. epidermidis ope
141	15	3.5	216	20	AAV17880	HCV NS4A-NS3 compl	214	9	2.1	26	12	AAW13621	Staphylococcus epi
142	15	3.5	216	20	AAV17881	HCV NS4A-NS3 compl	215	9	2.1	53	19	AAW36986	Amino acid sequenc
143	15	3.5	216	20	AAV17882	HCV NS4A-NS3 compl	216	9	2.1	54	19	AAW36990	Histidine tagged-h
144	15	3.5	216	20	AAV17883	HCV NS4A-NS3 compl	217	9	2.1	91	20	AAW09127	His-tagged Sonic h
145	15	3.5	216	20	AAV17884	HCV NS4A-NS3 compl	218	9	2.1	93	20	AAW09128	HIV TAT transduct
146	15	3.5	216	20	AAV17885	HCV NS4A-NS3 compl	219	9	2.1	134	22	AAW82144	B. stearothermophi
147	15	3.5	216	20	AAV17886	HCV NS4A-NS3 compl	220	9	2.1	139	23	ABW39568	Protein sequence o
148	15	3.5	216	20	AAV17887	HCV NS4A-NS3 compl	221	9	2.1	183	21	AAW84920	B. mori 10XHis-TRA
149	15	3.5	216	20	AAV17878	Hepatitis C virus	222	9	2.1	183	21	AAW70866	P. mori TRAS1 EN p
150	15	3.5	217	21	AAW44730	Cell factor deriva	223	9	2.1	192	21	AAW92865	FP505 protein cont
151	15	3.5	220	21	AAW88630	Human tyrosine kin	224	9	2.1	194	23	AAW51139	
152	15	3.5	239	21	AAW52366	His-OST311 protein	225	9	2.1	195	21	AAW11377	
153	15	3.5	248	23	ABW06951	Mercozole surfacis	226	9	2.1	254	18	AAW37699	
154	15	3.5	262	22	AAW37610	Bacillus anthracis	227	9	2.1	269	23	AAW49520	
155	15	3.5	288	23	AAW50707	Staphylococcus aur	228	9	2.1	269	23	AAW49541	
156	15	3.5	333	22	AAW02193		229	9	2.1	323	19	AAW71185	

230	9	2.1	332	22	AAE02191	Staphylococcus aur	303	7	1.6	10	22	AAB75181	Recombinant A33 an
231	9	2.1	364	21	AAG13011	Q-tagged FabH prot	304	7	1.6	11	20	AAB60389	Thrombin substrate
232	9	2.1	393	21	AAG68149	Streptomyces grise	305	7	1.6	10	20	AAW99048	Peptide 1f-trb-eps
233	9	2.1	393	22	AAW74844	S. griseus actinom	306	7	1.6	11	20	AAW99031	Peptide 1f-trb-eps
234	9	2.1	411	22	AAU00219	Bcl-Xl-DTR apoptos	307	7	1.6	12	21	AAH11895	H3SA mutant His-ta
235	9	2.1	423	19	AAW68098	Chlamydomonas rein	308	7	1.6	11	21	AAH11895	piII coat protein
236	9	2.1	448	19	AAW68399	Clostridium botuli	309	7	1.6	12	22	AAH11895	Human plasminogen
237	9	2.1	451	19	AAW68395	Clostridium botuli	310	7	1.6	12	22	AAU04934	Human plasminogen
238	9	2.1	451	19	AAW68398	Clostridium botuli	311	7	1.6	12	22	AAU04935	Sequence #1 for pa
239	9	2.1	452	19	AAW68396	Clostridium botuli	312	7	1.6	13	23	AAU77403	N-terminal His6 ta
240	9	2.1	462	17	AAW95009	Type A neurotoxin	313	7	1.6	14	21	AAU09365	Endostatin vector
241	9	2.1	462	19	AAW68390	Clostridium botuli	314	7	1.6	14	21	AAU09365	Cyclo-[Cys-His-Met
242	9	2.1	462	19	AAW68397	Clostridium botuli	315	7	1.6	14	21	AAH11893	T cell antigen rec
243	9	2.1	472	19	AAW68394	Clostridium botuli	316	7	1.6	14	21	AAH11893	pGEX-7HP-ro proteas
244	9	2.1	472	19	AAW68393	Clostridium botuli	317	7	1.6	17	20	AAW68977	N-terminal histidi
245	9	2.1	473	19	AAW68400	Clostridium botuli	318	7	1.6	17	20	AAW95000	Murine Toll-like r
246	9	2.1	491	23	ABW93536	Herbicidally activ	319	7	1.6	19	21	AAO21607	Protease Kex-2 cle
247	9	2.1	567	22	AAU00220	Bad-DTR apoptosis	320	7	1.6	19	21	AAW90748	Polypeptide fragme
248	9	2.1	600	23	AAE20482	Human protein for	321	7	1.6	21	21	AAW84796	Sequence of ident
249	8	1.9	8	21	AAE19475	Peptide derived fr	322	7	1.6	21	22	AAO5538	Human polypeptide
250	8	1.9	11	22	AAW63928	EGF-like domain of	323	7	1.6	23	23	AAO19984	Peptide of a prote
251	8	1.9	14	23	ABG67084	Streptavidin-bindi	324	7	1.6	24	22	AAW64594	OmpC-(6His)1. Syn
252	8	1.9	17	17	AAW03939	T2 peptide, synthe	325	7	1.6	24	22	AAO55865	Human polypeptide
253	8	1.9	17	20	ABW08913	TNF/hPI fusion pla	326	7	1.6	29	20	AAW43092	Tag sequence used
254	8	1.9	23	18	AAW5248	N-terminal peptide	327	7	1.6	32	23	ABW84805	Escherichia coli a
255	8	1.9	31	21	AAW15670	Alpha V beta 3 rec	328	7	1.6	37	22	AAW06798	Betalaphal/MBP-55
256	8	1.9	60	17	AAW90286	Nickel binding pro	329	7	1.6	37	22	AAW64595	OmpC-(6His)2. Syn
257	8	1.9	60	18	AAW20250	H. pylori cytoplas	330	7	1.6	41	21	AAW26109	H. contortus clone
258	8	1.9	61	21	AAW24617	H. pylori cytoplas	331	7	1.6	42	23	AAO19986	Protein of (double
259	8	1.9	61	21	AAW14344	Bovine pancreatic	332	7	1.6	42	23	AAO19986	Corticostatin. Sy
260	8	1.9	73	23	ABW80911	Murine eotaxin mat	333	7	1.6	45	14	AAU47506	Propionibacterium
261	8	1.9	78	18	AAW20756	H. pylori cytoplas	334	7	1.6	50	22	AAU47506	Propionibacterium
262	8	1.9	97	23	ABW80897	Murine eotaxin pol	335	7	1.6	54	23	ABP31575	Human kinase-like
263	8	1.9	98	23	ABW80898	Murine eotaxin pol	336	7	1.6	54	23	ABP31575	Human kinase-like
264	8	1.9	123	18	AAW28331	Staphylococcus aur	337	7	1.6	62	23	AAW85598	Ral2S-L985PEX pept
265	8	1.9	131	22	AAW81934	S. epidermidis ope	338	7	1.6	64	16	AAW82537	Hybrid IA beta cha
266	8	1.9	149	21	AAW08634	Arabidopsis thalia	339	7	1.6	65	16	AAW82535	Hybrid IE beta cha
267	8	1.9	155	21	AAW08633	Arabidopsis thalia	340	7	1.6	65	20	AAW91324	H6FXTN12 fusion pr
268	8	1.9	155	21	AAW43258	Arabidopsis thalia	341	7	1.6	68	22	AAW91324	Human immune/haema
269	8	1.9	158	21	AAW08632	Arabidopsis thalia	342	7	1.6	68	23	ABP02234	Human OREX protein
270	8	1.9	158	21	AAW43257	Arabidopsis thalia	343	7	1.6	77	22	AAU55128	Propionibacterium
271	8	1.9	159	21	AAW43256	Arabidopsis thalia	344	7	1.6	79	21	AAO2158	Human secreted pro
272	8	1.9	244	21	AAW93977	Synthetic green fl	345	7	1.6	80	22	AAU64509	Propionibacterium
273	8	1.9	244	21	AAW93978	Synthetic green fl	346	7	1.6	82	22	AAU64509	Human peptide #14
274	8	1.9	342	23	ABW33381	Human Trx_Wrl_B.	347	7	1.6	82	22	ABW27363	Peptide #15 encode
275	8	1.9	379	23	ABW54917	Lactococcus lactis	348	7	1.6	82	22	ABW32509	Peptide #15 encode
276	8	1.9	391	18	AAW23534	CDK inhibitory fus	349	7	1.6	82	22	ABW18015	Human brain expres
277	8	1.9	391	20	AAW95094	Human p27-p16 fusi	350	7	1.6	82	22	AAW53343	Human bone marrow
278	8	1.9	391	21	AAW97526	Human W3 protein s	351	7	1.6	82	22	AAW53343	Human bone marrow
279	8	1.9	391	21	AAW96041	Antiproliferative	352	7	1.6	82	22	AAW53343	Peptide #14 encode
280	8	1.9	391	21	AAW96068	Angiogenesis inhib	353	7	1.6	82	22	AAW53343	Peptide #15 encode
281	8	1.9	410	23	ABW33382	Human Trx_Wrl_A.	354	7	1.6	82	23	ABG33350	Human peptide enco
282	8	1.9	416	21	AAW18185	Plasmodium falcipa	355	7	1.6	89	22	AAW64597	OmpC-(6His)6. Syn
283	8	1.9	431	22	ABW69274	Drosophila melanog	356	7	1.6	91	23	ABW54331	Lactococcus lactis
284	8	1.9	544	23	ABW40264	Human peptide enco	357	7	1.6	91	23	ABW54331	Lactococcus lactis
285	8	1.9	591	23	ABG33383	Human Trx_Wrl. Ho	358	7	1.6	94	21	AAH13471	Protein encoded by
286	8	1.9	622	21	AAW97170	Human FGF-RI Extra	359	7	1.6	94	22	AAW83147	Protein encoded by
287	8	1.9	763	21	AAW38942	Arabidopsis thalia	360	7	1.6	94	23	ABW94118	Chlamydia protein
288	8	1.9	882	22	AAE07889	Modified clostridi	361	7	1.6	98	22	AAU42369	Propionibacterium
289	8	1.9	887	22	AAE07892	Modified clostridi	362	7	1.6	99	22	ABW3147	Peptide #6653 enco
290	8	1.9	974	22	ABW62110	Drosophila melanog	363	7	1.6	99	22	AAW59798	Human brain expres
291	8	1.9	1059	21	AAW93309	A manganese supero	364	7	1.6	99	22	AAW72385	Human bone marrow
292	8	1.9	1067	21	AAW93307	A manganese supero	365	7	1.6	99	22	AAW72385	Peptide #6676 enco
293	8	1.9	1070	21	AAW93308	A manganese supero	366	7	1.6	99	23	ABG42203	Human peptide enco
294	8	1.9	1084	21	AAW93312	A manganese supero	367	7	1.6	112	23	ABG42203	Human peptide enco
295	8	1.9	1092	21	AAW93310	A manganese supero	368	7	1.6	117	20	AAW21867	Amino acid sequenc
296	8	1.9	1095	21	AAW93311	A manganese supero	369	7	1.6	118	22	AAU31100	Novel human secret
297	7	1.6	7	9	AAW80401	Affinity peptide c	370	7	1.6	119	22	ABW30303	Peptide #2954 enco
298	7	1.6	7	23	AAE20476	Bovine ribonucleas	371	7	1.6	119	22	ABW35472	Peptide #2978 enco
299	7	1.6	8	15	AAW57097	N-terminal fragmen	372	7	1.6	119	22	ABW20911	Protein #2910 enco
300	7	1.6	8	16	AAW77463	GST-I-domain fusio	373	7	1.6	119	22	AAW56291	Human brain expres
301	7	1.6	8	16	AAW48167	Histidine tag to c	374	7	1.6	119	22	AAW68669	Human bone marrow
302	7	1.6	10	22	AAW98321	Recombinant A33 an	375	7	1.6	119	22	AAW16487	Peptide #2921 enco

376	7	1.6	119	22	AAW28977	Peptide #3014 enco	449	7	1.6	299	23	ABB75000	Ral2/C-terminal p
377	7	1.6	119	22	AMG04212	Peptide #2894 enco	450	7	1.6	299	23	AAE17587	Mycobacterium spec
378	7	1.6	119	23	ABG38250	Human peptide enco	451	7	1.6	302	21	RAY96926	M. tuberculosis an
379	7	1.6	120	20	RAY21857	Amino acid sequenc	452	7	1.6	304	22	AAU06902	Human /M. tubercul
380	7	1.6	121	20	RAY21858	Amino acid sequenc	453	7	1.6	304	22	AAU01257	Ral2-P775P-ORF3 co
381	7	1.6	128	18	AAW08427	Vector expressing	454	7	1.6	304	23	ABW95362	Ral2- P775P-ORF3 c
382	7	1.6	128	20	AAW30055	Amino acid sequenc	455	7	1.6	306	22	AAW74596	A fusion protein c
383	7	1.6	129	21	AAW13473	Protein encoded by	456	7	1.6	306	23	AAW74596	Antigenic fusion p
384	7	1.6	129	22	AAW83149	Protein encoded by	457	7	1.6	314	21	AAW33224	Eucalyptus grandis
385	7	1.6	129	23	ABW94239	Chlamydia protein	458	7	1.6	314	22	AAW11933	Human CG27 (or C86
386	7	1.6	130	18	AAW08185	TGF-beta fusion pr	459	7	1.6	314	23	AAW85589	Lung tumour protei
387	7	1.6	130	21	AAW34414	Arabidopsis thalia	460	7	1.6	320	21	AAW30655	Arabidopsis thalia
388	7	1.6	131	15	AAW63193	hEGF-St-IGF-1 fusi	461	7	1.6	320	21	AAW30655	Arabidopsis thalia
389	7	1.6	131	21	AAW58736	Breast and ovarian	462	7	1.6	321	20	AAW17300	Recombinant gp 90
390	7	1.6	133	22	AAW83268	Protein encoded by	463	7	1.6	324	22	ABG20509	Novel human diagno
391	7	1.6	133	23	ABW94239	Chlamydia protein	464	7	1.6	327	22	ABW69341	Drosophila melanog
392	7	1.6	134	18	AAW08187	TGF-beta fusion pr	465	7	1.6	330	22	AAW38918	C. trachomatis hyp
393	7	1.6	136	18	AAW08186	TGF-beta fusion pr	466	7	1.6	330	23	AAW74589	Antigenic fusion p
394	7	1.6	136	23	ABW06192	Protein library re	467	7	1.6	330	23	AAE17566	Mycobacterium sp.
395	7	1.6	137	22	AAW04232	Human polypeptide	468	7	1.6	330	23	AAE17567	Mycobacterium spec
396	7	1.6	137	22	AAW01885	M. tuberculosis an	469	7	1.6	331	20	AAW32060	Mycobacterium tube
397	7	1.6	158	22	AAW63415	Human breast cance	470	7	1.6	331	22	AAW66450	Pseudomonas aerugi
398	7	1.6	162	22	AAW01687	Gene 30 human secr	471	7	1.6	332	18	AAW32418	Mycobacterium tube
399	7	1.6	162	22	AAW82901	Histidine-tagged t	472	7	1.6	332	18	AAW32350	Mycobacterium tube
400	7	1.6	167	22	AAW64598	ParC-(6His)12. Sy	473	7	1.6	332	19	AAW81683	M. tuberculosis im
401	7	1.6	170	23	AAW11950	Parathyroid hormon	474	7	1.6	332	19	AAW64322	Mycobacterium tube
402	7	1.6	171	22	AAW13893	Human polypeptide	475	7	1.6	332	20	AAW39083	M. tuberculosis an
403	7	1.6	184	22	AAW30581	Amino acid sequenc	476	7	1.6	332	20	AAW38945	M. tuberculosis re
404	7	1.6	188	23	ABG33385	Human Wli-B. Homo	477	7	1.6	332	22	AAU01895	M. tuberculosis DP
405	7	1.6	191	21	AAW28398	Murine endostatin.	478	7	1.6	332	23	AAE17584	Mycobacterium spec
406	7	1.6	191	23	AAW77950	Amino acid sequenc	479	7	1.6	333	21	AAW41543	Arabidopsis thalia
407	7	1.6	196	18	AAW26746	S. carnosus nitrat	480	7	1.6	344	18	AAW22517	Heat-resistant ger
408	7	1.6	197	20	AAW94262	H6FXTN123 fusion p	481	7	1.6	344	22	AAW66427	Helicobacter pylor
409	7	1.6	202	21	AAW13474	Protein encoded by							

522	7	1.6	424	22	AAB48352	TRG-Ig fusion prot	595	7	1.6	672	13	AAR31216	penicillin binding
523	7	1.6	431	22	AAU55084	Propionibacterium	596	7	1.6	674	22	AAB66430	Pseudomonas aerugi
524	7	1.6	433	20	AAU32065	Mycobacterium tube	597	7	1.6	674	22	AAB66447	Pseudomonas aerugi
525	7	1.6	433	23	ABU94255	Chlamydia protein	598	7	1.6	676	22	AAB66428	Pseudomonas aerugi
526	7	1.6	433	23	AAU74594	Antigenic fusion p	599	7	1.6	676	22	AAB66445	Pseudomonas aerugi
527	7	1.6	435	23	ABG34126	Human clathrin coa	600	7	1.6	683	22	AAG83282	Chlamydia trachoma
528	7	1.6	445	19	AAW68391	Clostridium botuli	601	7	1.6	683	23	ABU94253	Chlamydia trachoma
529	7	1.6	447	21	AAU43728	Arabidopsis thalia	602	7	1.6	685	20	AAU00915	Human serum induci
530	7	1.6	449	22	AAU69872	Human prostate ser	603	7	1.6	685	20	AAW88432	Disease associated
531	7	1.6	449	22	AAU01227	P703P and PSA fusi	604	7	1.6	685	23	ABU61474	Human NF-kB activa
532	7	1.6	453	23	ABU95332	Human P703P/PSA fu	605	7	1.6	691	22	AAG83271	Chlamydia trachoma
533	7	1.6	453	18	AAU17784	FIV integrase-LexA	606	7	1.6	691	23	ABU94242	Chlamydia trachoma
534	7	1.6	459	16	AAW66729	Aromatic dihydrodi	607	7	1.6	700	22	AAG83279	Chlamydia trachoma
535	7	1.6	459	23	ABU92609	Herbicidally activ	608	7	1.6	700	23	ABU94250	Chlamydia trachoma
536	7	1.6	465	15	AAU63201	Wheat germ ACCase	609	7	1.6	702	22	ABU83844	Amino acid sequenc
537	7	1.6	466	22	AAU96071	Putative P. abyssi	610	7	1.6	706	22	ABU65428	Drosophila melanog
538	7	1.6	476	23	AAU76912	Human CRP2-4 /IgG	611	7	1.6	710	20	AAU32066	Mycobacterium tube
539	7	1.6	484	23	AAU76916	Human zcytorII/IgG	612	7	1.6	710	23	AAU17588	Mycobacterium spec
540	7	1.6	486	21	AAU28142	Arabidopsis thalia	613	7	1.6	715	20	AAU21860	Amino acid sequenc
541	7	1.6	487	22	AAU83280	Chlamydia trachoma	614	7	1.6	715	22	AAG83273	Chlamydia trachoma
542	7	1.6	487	23	ABU94251	Chlamydia trachoma	615	7	1.6	715	22	AAG83275	Chlamydia trachoma
543	7	1.6	499	18	AAU17783	FIV integrase-LexA	616	7	1.6	715	23	ABU94244	Chlamydia trachoma
544	7	1.6	507	21	AAU00158	SCD4-SCFV(17b) HIV	617	7	1.6	715	23	ABU94246	Chlamydia trachoma
545	7	1.6	511	23	AAU51137	Maltose binding pr	618	7	1.6	717	21	AAU28601	Arabidopsis thalia
546	7	1.6	517	16	AAU84619	Wheat acetyl-Coenz	619	7	1.6	717	22	ABU66734	Drosophila melanog
547	7	1.6	517	22	ABU61131	Drosophila melanog	620	7	1.6	723	23	AAU74595	Antigenic fusion p
548	7	1.6	518	22	AAU83276	Chlamydia trachoma	621	7	1.6	726	23	AAU74588	Antigenic fusion p
549	7	1.6	518	23	ABU94247	Chlamydia trachoma	622	7	1.6	729	20	AAU32059	Mycobacterium tube
550	7	1.6	521	21	AAU54022	Human pancreatic c	623	7	1.6	729	22	AAU22142	Ra12-H9-32A fusion
551	7	1.6	525	21	AAU13645	C. pneumoniae sero	624	7	1.6	729	23	AAU17572	Mycobacterium spec
552	7	1.6	525	22	AAU83213	Protein encoded by	625	7	1.6	729	23	AAU17573	Mycobacterium spec
553	7	1.6	525	23	ABU94184	Chlamydia protein	626	7	1.6	744	22	AAU01902	M. tuberculosis an
554	7	1.6	529	23	AAU74387	Breast tumour-spec	627	7	1.6	753	23	ABU41992	Human ovarian anti
555	7	1.6	543	22	AAU01905	M. tuberculosis an	628	7	1.6	768	23	ABU49385	Listeria monocytog
556	7	1.6	543	22	AAU66429	P. aeruginosa WpM	629	7	1.6	778	21	AAU13642	C. trachomatis pmp
557	7	1.6	543	22	AAU66446	Pseudomonas aerugi	630	7	1.6	778	22	AAU83210	Protein encoded by
558	7	1.6	553	22	ABU59793	Drosophila melanog	631	7	1.6	778	23	ABU94181	Chlamydia protein
559	7	1.6	554	17	AAU09252	K. pneumoniae diol	632	7	1.6	780	18	AAU11819	Candida albicans t
560	7	1.6	577	22	AAU22140	Ra12-WT1 fusion pr	633	7	1.6	785	23	ABU30430	Streptococcus poly
561	7	1.6	580	21	AAU82157	Pseudomonas putida	634	7	1.6	788	22	AAU01903	M. tuberculosis an
562	7	1.6	583	22	AAU83281	Chlamydia trachoma	635	7	1.6	815	22	AAU01904	M. tuberculosis an
563	7	1.6	583	23	ABU94252	Chlamydia trachoma	636	7	1.6	816	22	AAU03628	Group B Streptococ
564	7	1.6	585	22	AAU83277	Chlamydia trachoma	637	7	1.6	816	23	ABU25423	Streptococcus poly
565	7	1.6	585	23	ABU94248	Chlamydia trachoma	638	7	1.6	821	18	AAU16611	Candida albicans t
566	7	1.6	596	20	AAU32070	Mycobacterium tube	639	7	1.6	821	21	AAU13644	C. trachomatis pmp
567	7	1.6	596	22	AAU30220	Novel human secret	640	7	1.6	821	22	AAG83212	Protein encoded by
568	7	1.6	596	23	AAU17574	Mycobacterium spec	641	7	1.6	821	23	ABU94183	Chlamydia protein
569	7	1.6	597	22	AAU08231	Polypeptide encode	642	7	1.6	824	11	AAU04575	Derived amino acid
570	7	1.6	598	22	AAU38922	C. trachomatis CT6	643	7	1.6	825	12	AAU11254	Human IL-4 recepto
571	7	1.6	599	23	AAU74599	Antigenic fusion p	644	7	1.6	825	18	AAU13499	Human interleukin-
572	7	1.6	600	20	AAU32068	Mycobacterium tube	645	7	1.6	825	19	AAU60668	Human interleukin-
573	7	1.6	600	23	AAU74597	Antigenic fusion p	646	7	1.6	825	19	AAU48314	Human interleukin-
574	7	1.6	605	20	AAU21870	Amino acid sequenc	647	7	1.6	825	20	AAU93004	Human IL-4 recepto
575	7	1.6	609	20	AAU41279	Fusion protein con	648	7	1.6	825	20	AAU73470	Human interleukin-
576	7	1.6	619	22	AAU83270	Chlamydia trachoma	649	7	1.6	825	22	AAU38752	Human polypeptide
577	7	1.6	619	23	ABU94241	Chlamydia trachoma	650	7	1.6	825	22	AAU66970	Human IL4Ralpha
578	7	1.6	631	22	AAU83274	Chlamydia trachoma	651	7	1.6	825	22	AAU07612	Human interleukin-
579	7	1.6	631	23	ABU94245	Chlamydia trachoma	652	7	1.6	825	23	AAU77212	Human interleukin-
580	7	1.6	638	22	ABU58483	Drosophila melanog	653	7	1.6	825	23	AAU77214	Human interleukin-
581	7	1.6	646	20	AAU17892	HCV NS4A-NS3 compl	654	7	1.6	825	23	AAU78090	Human interleukin-
582	7	1.6	646	20	AAU17894	HCV NS4A-NS3 compl	655	7	1.6	825	23	AAU07427	Human interleukin-
583	7	1.6	646	20	AAU24950	HCV NS4A-NS3 compl	656	7	1.6	826	21	AAU95050	Candida albicans p
584	7	1.6	646	22	AAU83272	Chlamydia trachoma	657	7	1.6	826	22	AAU20493	Human interleukin-
585	7	1.6	646	23	ABU94243	Chlamydia trachoma	658	7	1.6	831	23	AAU77217	Human interleukin-
586	7	1.6	654	20	AAU21859	Amino acid sequenc	659	7	1.6	840	21	AAU93652	A mammalian solubi
587	7	1.6	654	22	AAU83278	Chlamydia trachoma	660	7	1.6	840	21	AAU70028	Soluble Interleuki
588	7	1.6	654	23	ABU94249	Chlamydia trachoma	661	7	1.6	856	20	AAU32064	Mycobacterium tube
589	7	1.6	660	22	AAU38921	C. trachomatis CT8	662	7	1.6	859	23	AAU74593	Antigenic fusion p
590	7	1.6	663	21	AAU52602	Arabidopsis thalia	663	7	1.6	861	22	ABU62534	Drosophila melanog
591	7	1.6	666	20	AAU21871	Amino acid sequenc	664	7	1.6	869	15	AAU56501	TATA-binding prote
592	7	1.6	666	22	ABU71578	Drosophila melanog	665	7	1.6	869	17	AAU60691	Human TATA-binding
593	7	1.6	667	20	AAU17893	HCV NS4A-NS3 compl	666	7	1.6	869	18	AAU25022	TATA-binding prote
594	7	1.6	667	20	AAU17891	HCV NS4A-NS3 compl	667	7	1.6	879	22	ABU25755	Novel human diagno

668	7	1.6	879	22	ABG28383	Novel human diagno	741	6	1.4	6	21	AAV99641	His6 tag used to c
669	7	1.6	888	23	ABB91325	Herbicidally activ	742	6	1.4	6	21	AAV87720	Thrombin recogniti
670	7	1.6	948	21	AAB13643	C. trachomatis pmp	743	6	1.4	6	21	AAV96204	Thrombin cleavage
671	7	1.6	948	22	AAG83211	Protein encoded by	744	6	1.4	6	21	AAV90140	Thrombin recogniti
672	7	1.6	948	23	ABB94182	Chlamydia protein	745	6	1.4	6	21	AAV90451	Trypsin substrate
673	7	1.6	950	21	AAG41703	Arabidopsis thalia	746	6	1.4	6	21	AAV81407	Thrombin cleavage
674	7	1.6	960	21	AAG31324	Arabidopsis thalia	747	6	1.4	6	21	AAV51517	Thrombin recogniti
675	7	1.6	982	23	AAU71855	Leishmania antigen	748	6	1.4	6	21	AAV77706	6-His peptide epit
676	7	1.6	989	23	ABB92703	Herbicidally activ	749	6	1.4	6	21	AAV77706	His-6 peptide SEQ
677	7	1.6	1024	22	ABB61798	Drosophila melanog	750	6	1.4	6	21	AAV78346	His-6 peptide SEQ
678	7	1.6	1068	20	AAV30048	Pancreatic eukaryo	751	6	1.4	6	21	AAV78351	His tag. Unidentl
679	7	1.6	1079	22	AAV74830	Prostate tumour an	752	6	1.4	6	22	ABB76820	Peptide tag 1. Sy
680	7	1.6	1115	20	AAV30047	Pancreatic eukaryo	753	6	1.4	6	22	AAV52173	Epitope tag #7 use
681	7	1.6	1115	23	AAU81285	EIF-2alpha kinase.	754	6	1.4	6	22	AAE13079	FSH alpha or beta
682	7	1.6	1161	22	ABB70419	Drosophila melanog	755	6	1.4	6	22	AAE12713	Hexahis tag used t
683	7	1.6	1186	22	ABG24444	Drosophila melanog	756	6	1.4	6	22	AAE12713	His tag used for r
684	7	1.6	1225	22	ABG24444	Novel human diagno	757	6	1.4	6	22	AAE08742	Poly-His tag, to g
685	7	1.6	1255	20	AAV05944	Thermophilus therm	758	6	1.4	6	22	AAE08742	6-His epitope tag.
686	7	1.6	1284	22	AAV05944	Drosophila melanog	759	6	1.4	6	22	AAE08742	Amino acid sequenc
687	7	1.6	1284	23	ABB08376	D. melanogaster CG	760	6	1.4	6	22	AAE08006	His tag useful as
688	7	1.6	1313	14	AAR36781	A3 maize ACCase.	761	6	1.4	6	22	AAE08006	Metal capturing pr
689	7	1.6	1427	23	AAU71857	Leishmania MAPS1A-	762	6	1.4	6	22	AAU02090	His-tag peptide.
690	7	1.6	1465	22	ABB64500	Drosophila melanog	763	6	1.4	6	22	AAU02090	6-His epitope used
691	7	1.6	1534	22	ABB58402	Drosophila melanog	764	6	1.4	6	22	AAE02682	C-terminal His tag
692	7	1.6	1561	22	ABB56445	Drosophila melanog	765	6	1.4	6	22	AAE01997	Poly-His tag used
693	7	1.6	1641	22	AAU71856	Drosophila melanog	766	6	1.4	6	22	AAU74834	MSAI gp42 C-termin
694	7	1.6	1807	22	AAE05697	Recombinant protei	767	6	1.4	6	22	AAU00163	His Tag for purifi
695	7	1.6	1895	22	ABB70088	Recombinant protei	768	6	1.4	6	22	AAU00163	Nascent protein de
696	7	1.6	1911	22	ABB61038	Drosophila melanog	769	6	1.4	6	22	AAE35439	6-His epitope. Un
697	7	1.6	2028	22	AAE05698	Recombinant protei	770	6	1.4	6	22	AAE35439	Transdominant effe
698	7	1.6	2091	22	ABB70224	Drosophila melanog	771	6	1.4	6	22	AAE35439	6-His peptide epit
699	7	1.6	2280	22	ABB61650	Drosophila melanog	772	6	1.4	6	22	AAE35439	Human IP3 binding
700	7	1.6	2311	23	AAU11820	Wheat plastid acet	773	6	1.4	6	22	AAE35439	Human IP3 binding
701	7	1.6	2325	17	AAU05590	Maize acetyl CoA c	774	6	1.4	6	22	AAE35439	His6 tag used in t
702	7	1.6	2325	21	AAW56736	Maize ACCase enzym	775	6	1.4	6	23	AAU97024	Cattle enteropepti
703	7	1.6	2325	21	AAV44687	Maize acetyl CoA c	776	6	1.4	6	23	AAU97024	AAV-helper plasmid
704	7	1.6	3080	22	ABB64877	Drosophila melanog	777	6	1.4	6	23	AAU97024	Thrombin cleavage
705	7	1.6	5533	22	ABB55772	Drosophila melanog	778	6	1.4	6	23	AAU97024	Thrombin cleavage
706	7	1.6	5560	22	ABB71160	Drosophila melanog	779	6	1.4	6	23	AAU97024	KpnI/Sali His-tag
707	7	1.6	6	15	AAW05053	Factor Xa cleavage	780	6	1.4	6	23	AAU97024	His6 tag used in th
708	6	1.4	6	16	AAE86305	Thrombin cleavage	781	6	1.4	6	23	AAU97024	His6 tag used in th
709	6	1.4	6	16	AAE71098	Prod. of primer to	782	6	1.4	6	23	AAE71098	His6 tag used in th
710	6	1.4	6	16	AAE71098	proteolytic cleava	783	6	1.4	6	23	AAE71098	Thrombin cleavage
711	6	1.4	6	16	AAE77462	GST-I-domain fusio	784	6	1.4	6	23	AAE71098	Epitope tag #5 fus
712	6	1.4	6	18	AAW30085	Human calcitonin g	785	6	1.4	6	23	AAU10567	Purification pepti
713	6	1.4	6	18	AAW08170	Thrombin proteinase	786	6	1.4	6	23	AAU10567	Endoproteinase Xa
714	6	1.4	6	18	AAW18225	Purification tag o	787	6	1.4	6	23	AAU10567	Human cardiac tro
715	6	1.4	6	18	AAW19161	Isoelectric point	788	6	1.4	6	23	AAU10567	Peptide insert #3
716	6	1.4	6	19	AAW63024	Hexa-histidine tag	789	6	1.4	6	23	AAU10567	Tag added to C-ter
717	6	1.4	6	19	AAW69961	Poly-His tag for c	790	6	1.4	6	23	AAU10567	Extension peptide
718	6	1.4	6	19	AAW68297	Poly-His tag for c	791	6	1.4	6	23	AAU10567	Trypsinogen activa
719	6	1.4	6	19	AAW51545	Pre-sequence which	792	6	1.4	6	23	AAU10567	Thrombin cleavage
720	6	1.4	6	19	AAW44011	Poly-histidine pep	793	6	1.4	6	23	AAU10567	Polyhistidine tail
721	6	1.4	6	20	AAV41292	Thrombin cleavage	794	6	1.4	6	23	AAU10567	C-terminal metal b
722	6	1.4	6	20	AAV49221	Nickel ion binding	795	6	1.4	6	23	AAU10567	A. fumigatus aller
723	6	1.4	6	20	AAV50711	Single chain antib	796	6	1.4	6	23	AAU10567	Neutrophil-activat
724	6	1.4	6	20	AAV33592	VH-VL domain linke	797	6	1.4	6	23	AAU10567	Hexahistidine tag
725	6	1.4	6	20	AAV33835	Amino acid sequenc	798	6	1.4	6	23	AAU10567	C. tetani tetanus
726	6	1.4	6	20	AAV29844	Thrombin cleavage	799	6	1.4	6	23	AAU10567	Tag fused to MN pr
727	6	1.4	6	20	AAV31223	Thrombin excision	800	6	1.4	6	23	AAU10567	Polyhistidine affi
728	6	1.4	6	20	AAV06469	Epitope tag. Synt	801	6	1.4	6	23	AAU10567	Consensus thrombin
729	6	1.4	6	20	AAV17447	Major outer membra	802	6	1.4	6	23	AAU10567	Peptide tag 2. Sy
730	6	1.4	6	20	AAV09544	Human TAK1 6xHis p	803	6	1.4	6	23	AAU10567	His tag used for a
731	6	1.4	6	20	AAW99047	Peptide If-Xa-epsi	804	6	1.4	6	23	AAU10567	FSH alpha or beta
732	6	1.4	6	20	AAW99030	Peptide If-Xa-epsi	805	6	1.4	6	23	AAU10567	Tag #2, to facilit
733	6	1.4	6	20	AAW94257	Bovine restriction	806	6	1.4	6	23	AAU10567	Tag #1 for human
734	6	1.4	6	20	AAW84159	Peptide comprising	807	6	1.4	6	23	AAU10567	Peptide tag. Unid
735	6	1.4	6	20	AAW84203	Peptide comprising	808	6	1.4	6	23	AAU10567	Peptide insert #4.
736	6	1.4	6	21	AAW808438	Histidine tag rela	809	6	1.4	6	23	AAU10567	His Tag (2) for pu
737	6	1.4	6	21	AAW35702	Peptide 1 used in	810	6	1.4	6	23	AAU10567	Peptide used in re
738	6	1.4	6	21	AAW10788	B. subtilis lumazi	811	6	1.4	6	23	AAU10567	Brevibacillus chos
739	6	1.4	6	21	AAW28983	Peptide Y2X thromb	812	6	1.4	6	23	AAU10567	His6tag peptide se
740	6	1.4	6	21	AAV96951	Polyhistidine tag	813	6	1.4	6	23	AAU10567	

814	6	1.4	8	23	AAU99814	Membrane associate	887	6	1.4	11	23	ABB76132	Modified polyhisti
815	6	1.4	8	23	AAU96775	Divinyl ether synt	888	6	1.4	11	23	AAE20416	His6Gly4Cys spacer
816	6	1.4	8	23	AAU78663	His tag sequence.	889	6	1.4	11	23	AAAM50625	Polyhistidine tag.
817	6	1.4	8	23	AAU77821	C terminal extensio	890	6	1.4	12	11	AAK07064	Transcript of plas
818	6	1.4	8	23	AAU77832	His tag peptide.	891	6	1.4	12	15	AAK60140	pGEX-27k protease
819	6	1.4	8	23	AAE16245	Polyhistidine tag,	892	6	1.4	12	16	AAK66449	GST-hippuricase fu
820	6	1.4	9	17	AAU06112	Variant adenovirus	893	6	1.4	12	18	AAW10571	Factor Xa recognit
821	6	1.4	9	17	AAU97378	Penta-histidine pe	894	6	1.4	12	19	AAW61549	Endoproteinase Xa
822	6	1.4	9	19	AAW60080	Homo sapiens RH ep	895	6	1.4	12	20	AAW16775	Histidine tag used
823	6	1.4	9	20	AAV29934	FLAG peptide. Syn	896	6	1.4	12	20	AAW94420	Plasmidium falci
824	6	1.4	9	20	AAV28306	Amino acid sequenc	897	6	1.4	12	20	AAW82993	Human fibronectin
825	6	1.4	9	20	AAW90139	B7.2-6His tag fusi	898	6	1.4	12	21	AAV95878	Hexahistidine tag
826	6	1.4	9	20	AAW82990	Human fibronectin	899	6	1.4	12	22	AAW64365	Peptide insert #5.
827	6	1.4	9	21	AAW15187	His-containing pep	900	6	1.4	12	22	AAW85050	Peptide tag for re
828	6	1.4	9	21	AAW79558	Peptide tag 6xHis.	901	6	1.4	12	22	AAW83199	FLAG epitope tag.
829	6	1.4	9	21	AAV53258	Human neurotactin	902	6	1.4	12	22	AAW72438	His tag. Unidenti
830	6	1.4	9	21	AAV44463	Gly(His)8 tag, gh	903	6	1.4	12	23	AAU96774	Divinyl ether synt
831	6	1.4	9	22	AAE13114	C-terminal tag of	904	6	1.4	12	23	AAU96716	Molecular marker f
832	6	1.4	9	22	AAE82652	His-tag. Syntheti	905	6	1.4	12	23	AAU19655	Arabidopsis thalia
833	6	1.4	9	22	AAE64361	Peptide insert #1.	906	6	1.4	12	23	AAW05983	Monoclonal antibod
834	6	1.4	9	22	AAW99760	Rhesus D antibody	907	6	1.4	13	16	AAW72792	Hexa-histidine lea
835	6	1.4	9	22	AAV97661	Influenza haemaggl	908	6	1.4	13	19	AAW69482	Partial sequence o
836	6	1.4	9	23	AAU09959	Nonapeptide encodi	909	6	1.4	13	19	AAW61548	Endoproteinase Xa
837	6	1.4	10	14	AAK39773	First type QE- C-t	910	6	1.4	13	20	AAV17524	P. carinii tyrosyl
838	6	1.4	10	15	AAK73684	Metal-affinity dec	911	6	1.4	13	20	AAV05816	Ngn-toxic modified
839	6	1.4	10	16	AAW74672	Polyhistidine moti	912	6	1.4	13	21	AAW10782	B. subtilis lumazi
840	6	1.4	10	16	AAW77469	I-domain N-termina	913	6	1.4	13	21	AAW14343	Coriolus versicolo
841	6	1.4	10	18	AAW31496	Human DNase II pro	914	6	1.4	13	21	AAV56262	Maedi-Visna virus
842	6	1.4	10	18	AAW30084	Human calcitonin g	915	6	1.4	13	22	AAE13116	C-terminal tag of
843	6	1.4	10	19	AAW72331	Protein allergen I	916	6	1.4	13	22	AAU09362	N-terminal HMK-His
844	6	1.4	10	19	AAW56089	Human monocyte che	917	6	1.4	13	22	AAU09363	N-terminal HMK-His
845	6	1.4	10	20	AAV43534	Leishmania mexican	918	6	1.4	13	22	AAU09364	N-terminal HMK-His
846	6	1.4	10	20	AAV50501	Dermatophagoides s	919	6	1.4	13	22	AAE12152	Oligopeptide used
847	6	1.4	10	20	AAV06610	Hexahistidine tag	920	6	1.4	13	22	AAE80817	His tag peptide #1
848	6	1.4	10	20	AAW								

960 6 1.4 15 22 AAB66370 Hepatitis C virus
961 6 1.4 15 23 AAE20475 Human pancreatic r
962 6 1.4 15 23 AAE19656 Arabidopsis thalia
963 6 1.4 15 23 AAB77664 His tag peptide.
964 6 1.4 15 23 ABB05375 15 amino acid His
965 6 1.4 16 14 AAR44552 Feline leukemia v
966 6 1.4 16 16 AAW11923 Residues 146-161 o
967 6 1.4 16 16 AAR69813 Cleavage site reco
968 6 1.4 16 18 AAW8014 Plasmodium falcipa
969 6 1.4 16 20 AAY06321 His-c-myc epitope
970 6 1.4 16 21 AAY51161 Modified VPI pepti
971 6 1.4 16 22 AAU27552 Tag sequence used
972 6 1.4 16 22 AAY72938 N-terminal of huma
973 6 1.4 16 22 AAY97659 Influenza haemaggl
974 6 1.4 16 22 AAB66989 Peptide: SEQ ID 16
975 6 1.4 16 23 ABB08039 Amino acid fragmen
976 6 1.4 16 23 AAU77202 Polyasparagine, he
977 6 1.4 16 23 AAM47310 GMT-JMT fusion pro
978 6 1.4 17 13 AAR25383 Junction sequence
979 6 1.4 17 14 AAR37708 Delta14 Ser17 hcNT
980 6 1.4 17 17 AAR97389 CC49 VH-spacer-PLA
981 6 1.4 17 21 AAB26262 N-terminal fusion
982 6 1.4 17 22 AAG99608 ERA binding domain
983 6 1.4 18 15 AAR60253 Vector-encoded NF-
984 6 1.4 18 16 AAR78152 Synthetic leader t
985 6 1.4 18 16 AAR69814 Site recognised En
986 6 1.4 18 17 AAR95870 CR2 receptor ligand
987 6 1.4 18 21 AAB11047 Staphylococcal sur
988 6 1.4 18 21 AAB10856 pSectag-VEGF prote
989 6 1.4 18 22 ABB28438 Peptide #1089 enco
990 6 1.4 18 22 ABB31815 Peptide #4466 enco
991 6 1.4 18 22 ABB33620 Peptide #1126 enco
992 6 1.4 18 22 ABB37046 Peptide #4552 enco
993 6 1.4 18 22 ABB19075 Protein #1074 enco
994 6 1.4 18 22 ABB22361 Protein #4360 enco
995 6 1.4 18 22 AAB85750 Mouse sonic hedgeh
996 6 1.4 18 22 AAM54393 Human brain expres
997 6 1.4 18 22 AAM57773 Human brain expres
998 6 1.4 18 22 AAM66794 Human bone marrow
999 6 1.4 18 22 AAM70188 Human bone marrow
1000 6 1.4 18 22 AAM14660 Peptide #1094 enco

ALIGNMENTS

RESULT 1
AAB83926
ID AAB83926 standard; Protein; 375 AA.
XX AAB83926;
XX 23-JUL-2001 (first entry)
XX A major merozoite surface protein-1 fragment of 42kda.
DE A major merozoite surface protein-1; MSPI-42; melittin signal peptide;
XX Major merozoite surface protein-1; MSPI-42; melittin signal peptide;
XX malaria vaccine.
XX Plasmodium falciparum.
XX WC2001134188-A1
XX 17-MAY-2001.
XX 09-NOV-2000; 2000WO-US31064.
XX 12-NOV-1999; 99US-0165178.
XX 01-DEC-1999; 99US-0168327.
XX 22-AUG-2000; 2000US-0226861.
XX (UYHA-) UNIV HAWAII.
XX (UYCH-) UNIV CHINESE HONG KONG.

NEW
102(e)

PA (QUEE-) QUEEN EMMA FOUND.
XX Hui GSN, Lap-Yin P, Ho WKK;
XX WPI; 2001-335879/35.
XX N-PSDB; AAF89840.
XX Producing malaria vaccine, useful for treatment or prevention of all
XX forms of malaria in humans, by expressing immunogenic merozoite protein
XX fragment in a baculovirus system
XX Example 3; Page 87-88; 95pp; English.
XX The present sequence represents a major merozoite surface protein-1
XX C-terminal fragment of 42kDa (MSPI-42). This fragment is linked to a
XX melittin signal peptide, and then expressed in a in a
XX silkworm/baculovirus system. The protein is used to prepare a
XX malaria vaccine, which is used to treat or prevent malaria, caused by
XX any of the four species of Plasmodium that infect humans.
XX Sequence 375 AA;

Query Match 87.0%; Score 375; DB 22; Length 375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 57 MAISVTMDNILSGFENEYDVIYKPLAGVYRSLKKQIEKNFTFNLDILNSRLKRRK 116
Db 1 MAISVTMDNILSGFENEYDVIYKPLAGVYRSLKKQIEKNFTFNLDILNSRLKRRK 60
Qy 117 YFLVDLESDLAQFKHISSEYIIEEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGI 176
Db 61 YFLVDLESDLAQFKHISSEYIIEEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGI 120
Qy 177 SYEYKVLAKYKDDLESIKKVIKEKEKPPSPPTTPPSPAKTDEOKKESKFLPFTNIET 236
Db 121 SYEYKVLAKYKDDLESIKKVIKEKEKPPSPPTTPPSPAKTDEOKKESKFLPFTNIET 180
Qy 237 LYNNLVNKIDYDYLNLKAKINDCNVKEDEAHVITKLSDLKAIDDKIDLFKNPYDFAIK 296
Db 181 LYNNLVNKIDYDYLNLKAKINDCNVKEDEAHVITKLSDLKAIDDKIDLFKNPYDFAIK 240
Qy 297 KLINDTKKMDLGLSTGLVONFPNTIISKLEGGKFDMDLNIHQCVKQCPENSGCF 356
Db 241 KLINDTKKMDLGLSTGLVONFPNTIISKLEGGKFDMDLNIHQCVKQCPENSGCF 300
Qy 357 RHLDEREECKLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSRKKITCECT 416
Db 301 RHLDEREECKLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSRKKITCECT 360
Qy 417 KPDSYPLFDGIFCSS 431
Db 361 KPDSYPLFDGIFCSS 375

RESULT 2
AAW36103
ID AAW36103 standard; Protein; 116 AA.
XX AAW36103;
XX 25-MAR-1998 (first entry)
XX PfMSPI(pl9)A protein sequence.
XX Plasmodium vivax; merozoite surface protein; MSPI; pl9;
XX Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX Plasmodium falciparum.
XX Synthetic.
XX Key Location/Qualifiers
XX Region 1..95


```
FT FT /note= "amino acids derived from P. falciparum MSP1 p19  
FT FT fragment"  
FT FT 96..116  
XX /note= "glycosylphosphatidylinositol anchoring sequence"  
XX WO9730158-A2.  
XX 21-AUG-1997.  
XX (INSP ) INST PASTEUR.  
XX (UWNY ) UNIV NEW YORK STATE.  
XX 14-FEB-1997; 97WO-FR00290.  
XX 14-FEB-1996; 96FR-0001822.  
XX (INSP ) INST PASTEUR.  
XX (UWNY ) UNIV NEW YORK STATE.  
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;  
XX Roth C;  
XX WPI; 1997-425033/39.  
XX N-PSDB; AAT94550.  
XX Recombinant protein containing the merozoite surface protein-1 p19  
XX fragment - useful in anti-malarial vaccines, diagnosis and protein  
XX purification  
XX Disclosure; Fig 1B; 85pp; French.  
XX This is the amino acid sequence of a recombinant protein comprising  
XX amino acids 1613-1705 of the Plasmodium falciparum merozoite surface  
XX protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a  
XX glycosylphosphatidylinositol membrane anchoring sequence. p19 is the  
XX C-terminal fragment of the 42 kD MSP1 from Plasmodium species.  
XX The recombinant protein can be used for the production of anti-malarial  
XX vaccines, where the p19 fragment provides a high level of protective  
XX immunity since it includes epitopes not presented in the p42 fragment.  
XX SQ Sequence 116 AA;  
Query Match 13.9%; Score 60; DB 18; Length 116;  
Best Local Similarity 100.0%; Pred. No. 3.7e-52;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 338 NTSQHCVKQCPENSGCFRHLDERECKLLNKKQEGDKCVENPNTCNENNGGDADA 397  
DB 3 NTSQHCVKQCPENSGCFRHLDERECKLLNKKQEGDKCVENPNTCNENNGGDADA 62  
RESULT 3  
AAW22592  
ID AAW22592 standard; Protein; 116 AA.  
XX AC AAW22592;  
XX 25-MAR-1998 (first entry)  
XX PfMSP1(p19)A protein sequence.  
XX Plasmodium vivax; merozoite surface protein; MSP1; p19;  
XX Plasmodium falciparum; malaria; vaccine; immunity; epitope.  
XX Plasmodium falciparum.  
XX Synthetic.  
XX Key Location/Qualifiers  
XX FT Region 1..95  
XX FT /note= "amino acids derived from P. falciparum MSP1 p19  
XX FT fragment"  
XX FT 96..116  
XX FT /note= "glycosylphosphatidylinositol anchoring sequence"  
XX WO9730159-A2.  
XX
```

```
PD 21-AUG-1997.  
XX 14-FEB-1997; 97WO-FR00291.  
XX 14-FEB-1996; 96FR-0001821.  
XX (INSP ) INST PASTEUR.  
XX (UWNY ) UNIV NEW YORK STATE.  
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;  
XX Roth C;  
XX WPI; 1997-425034/39.  
XX P-PSDB; AAW22592.  
XX Recombinant protein containing Plasmodium merozoite surface.  
XX protein-1 p42 fragment - useful in antimalarial vaccines, also new  
XX antibodies for diagnosis and protein purification  
XX Disclosure; Fig 1B; 85pp; French.  
XX This is the amino acid sequence of a recombinant protein comprising  
XX amino acids 1613-1705 of the Plasmodium falciparum merozoite surface  
XX protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a  
XX glycosylphosphatidylinositol membrane anchoring sequence. p19 is the  
XX C-terminal fragment of the 42 kD MSP1 from Plasmodium species.  
XX The recombinant protein can be used for the production of anti-malarial  
XX vaccines, where the p19 fragment provides a high level of protective  
XX immunity since it includes epitopes not presented in the p42 fragment.  
XX SQ Sequence 116 AA;  
Query Match 13.9%; Score 60; DB 18; Length 116;  
Best Local Similarity 100.0%; Pred. No. 3.7e-52;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 338 NTSQHCVKQCPENSGCFRHLDERECKLLNKKQEGDKCVENPNTCNENNGGDADA 397  
DB 3 NTSQHCVKQCPENSGCFRHLDERECKLLNKKQEGDKCVENPNTCNENNGGDADA 62  
RESULT 4  
AAW22593  
ID AAW22593 standard; Protein; 127 AA.  
XX AC AAW22593;  
XX 25-MAR-1998 (first entry)  
XX PfMSP1(p19)S protein sequence.  
XX Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;  
XX Plasmodium falciparum; malaria; vaccine; immunity; epitope.  
XX Chimeric - Plasmodium vivax.  
XX Chimeric - Plasmodium falciparum.  
XX Key Location/Qualifiers  
XX FT Peptide 1..19  
XX FT /note= "signal peptide"  
XX FT Protein 20..127  
XX FT /note= "mature protein"  
XX FT Region 1..32  
XX FT /note= "derived from P. vivax MSP1"  
XX FT Region 33..34  
XX FT /note= "encoded by restriction enzyme sequence used to  
XX FT create the chimeric sequence"  
XX FT Region 35..127  
XX FT /note= "derived from P. falciparum C-terminal p19  
XX FT fragment of MSP1"  
XX WO9730159-A2.  
XX
```

PD 21-AUG-1997.
XX
XX 14-FEB-1997; 97WO-FR00291.
XX
PR 14-FEB-1996; 96FR-0001821.
XX
XX (INSP) INST PASTEUR.
PA (UYN) UNIV NEW YORK STATE.
XX
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
PI Roth C;
PI
XX
DR WPI; 1997-425034/39.
DR P-PSDB; AAW22592.
XX
XX Recombinant protein containing Plasmodium merozoite surface
PT protein-1 p42 fragment - useful in antimalarial vaccines, also new
PT antibodies for diagnosis and protein purification
XX
XX Disclosure; Fig 1C; 85pp; French.
XX
XX This is the amino acid sequence of a chimeric protein comprising amino
CC acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSp1)
CC linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
CC C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC The recombinant protein can be used for the production of anti-malarial
CC vaccines, where the p19 fragment provides a high level of protective
CC immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ Sequence 127 AA;
Query Match 13.9%; Score 60; DB 18; Length 127;
Best Local Similarity 100.0%; Pred. No. 4e-52;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 338 NISQHCVKKQCPNSGCFRHLDERECKLLNYKQBGDKCVENPNTCTNENGGCDADA 397
DB 35 NISQHCVKKQCPNSGCFRHLDERECKLLNYKQBGDKCVENPNTCTNENGGCDADA 94
RESULT 5
AAW36102
ID AAW36102 standard; Protein; 127 AA.
XX
AC AAW36102;
XX
XX 25-MAR-1998 (first entry)
XX
XX PFMSP1(p19)S protein sequence.
XX
XX Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
XX Chimeric - Plasmodium vivax.
OS Chimeric - Plasmodium falciparum.
XX
XX Key Location/Qualifiers
FH Peptide 1..19 /note= "signal peptide"
FT Protein 20..127 /note= "mature protein"
FT Region 1..32 /note= "derived from P. vivax MSP1"
FT Region 33..34 /note= "encoded by restriction enzyme sequence used to
FT Region 35..127 create the chimeric sequence"
FT /note= "derived from P. falciparum C-terminal p19
FT fragment of MSP1"
XX
PN WO9730158-A2.
XX

PD 21-AUG-1997.
XX
XX 14-FEB-1997; 97WO-FR00290.
XX
PR 14-FEB-1996; 96FR-0001822.
XX
XX (INSP) INST PASTEUR.
PA (UYN) UNIV NEW YORK STATE.
XX
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
PI Roth C;
PI
XX
DR WPI; 1997-425033/39.
DR N-PSDB; AAT94549.
XX
XX Recombinant protein containing the merozoite surface protein-1 p19
PT fragment - useful in anti-malarial vaccines, diagnosis and protein
PT purification
XX
XX Disclosure; Fig 1C; 85pp; French.
XX
XX This is the amino acid sequence of a chimeric protein comprising amino
CC acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSp1)
CC linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
CC C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC The recombinant protein can be used for the production of anti-malarial
CC vaccines, where the p19 fragment provides a high level of protective
CC immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ Sequence 127 AA;
Query Match 13.9%; Score 60; DB 18; Length 127;
Best Local Similarity 100.0%; Pred. No. 4e-52;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 338 NISQHCVKKQCPNSGCFRHLDERECKLLNYKQBGDKCVENPNTCTNENGGCDADA 397
DB 35 NISQHCVKKQCPNSGCFRHLDERECKLLNYKQBGDKCVENPNTCTNENGGCDADA 94
RESULT 6
AAR41357
ID AAR41357 standard; peptide; 54 AA.
XX
AC AAR41357;
XX
XX 04-MAR-1994 (first entry)
XX
XX MSP1EGF2B EGF2-like domain variant.
XX
XX Epidermal growth factor 1; merozoite surface protein 1; malaria;
KW vaccine.
XX
XX Plasmodium yoelii.
OS
XX Key Location/Qualifiers
FH Cleavage-site 1 /note= "introduced to facilitate cleavage
FT from recombinant protein"
FT
XX
PN WO9317107-A.
XX
XX 02-SEP-1993.
XX
XX 22-FEB-1993; 93WO-GB00367.
XX
XX 22-FEB-1992; 92GB-0003821.
XX
XX (MEDI-) MEDICAL RES COUNCIL.
XX
XX Blackman MJ, Chappel JA, Holder AA;
XX

DR WPI; 1993-288413/36.
XX Allelic variants of epidermal growth factor 1- or 2-like domains - of
PT merozoite surface protein 1, produced recombinantly for malaria
PT vaccines
XX
XX
PS Claim 2; Fig 2b; 35pp; English.
XX The sequence is that of an allelic variant of a merozoite surface
CC protein-1 epidermal growth factor (EGF) 2-like domain. It may be
CC used alone or as part of a fusion protein of EGF-1-like and
CC EGF-2-like domains in vaccines against malaria. When expressed
CC recombinantly it is produced in a form indistinguishable from
CC that in the native protein.
XX
XX
SQ Sequence 54 AA;
Query Match 11.8%; Score 51; DB 14; Length 54;
Best Local Similarity 100.0%; Pred. No. 2.1e-43;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 381 NPNTCNENGGCDATCTEEDSGSRKKITCECTKPSYPLFDGIFCSS 431
DB 2 NPNTCNENGGCDATCTEEDSGSRKKITCECTKPSYPLFDGIFCSS 52
RESULT 7
AAB26110
ID AAB26110 standard; Protein; 76 AA.
XX
AC AAB26110;
XX
XX 30-JAN-2001 (first entry)
XX
XX H. contortus clone-65e vector pET30a cloning junction #2.
XX
XX Nematode; parasite; helminth; sheep; goat; stomach; vaccine.
XX
XX Synthetic.
XX
XX WO200056763-A1.
XX
XX 28-SEP-2000.
XX
XX 16-MAR-2000; 2000WO-AU00210.
XX
XX 18-MAR-1999; 99AU-0009297.
XX
XX (NOVS) NOVARTIS AG.
XX
XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.
XX
XX Savin KW, Cook VR, Chen Y, Sexton JL, Apos E, Wilson LR;
PI Griffiths TM, Newton SE;
XX
XX WPI; 2000-594573/56.
DR N-PSDB; AAA94064.
XX
XX New Haemonchus contortus polypeptide for inducing a protective effect
PT against a helminth by controlling helminth infection, growth, viability
PT and/or egg fecundity and for ameliorating the symptoms of helminth
PT infection -
XX
XX Disclosure; Fig 8; 94pp; English.
XX
XX The present sequence is the translated sequence found at the cloning
CC junction in the pET30a vector of nematode Haemonchus contortus
CC clone 65e. This clone encodes a novel protein. This organism is a
CC parasite found in the stomach of its host (sheep and goats in
CC particular). The protein can be used in a vaccine against other
CC helminths, as well as Haemonchus contortus, including trematodes,
CC cestodes, nematodes and acanthocephala. These are all capable of causing
CC severe illness in their hosts, which include sheep, pigs, goats, cattle,
CC horses, donkeys, dogs, cats, guinea pigs and cage-birds, along with

CC humans. Antibodies to the protein can be used to diagnose infection.
CC Note: This sequence is stated as being the same as that shown in SEQ ID
CC NO: 6 of the specification (see AAB26109). However, this sequence is
CC shorter than the one shown here.
XX
XX
SQ Sequence 76 AA;
Query Match 11.6%; Score 50; DB 21; Length 76;
Best Local Similarity 100.0%; Pred. No. 3e-42;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDGDDDDKAMADIGS 50
DB 1 MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDGDDDDKAMADIGS 50
RESULT 8
AAB85253
ID AAB85253 standard; Protein; 167 AA.
XX
AC AAB85253;
XX
XX 07-SEP-2001 (first entry)
XX
XX Thioredoxin functional fragment.
DE
XX
XX Thioredoxin; porphobilinogen synthase; T-PPS; enzyme; herbicide;
XX delta-aminolevulinic acid; plant growth inhibitor.
XX
XX Lycopersicon esculentum.
XX
XX WO200146446-A1.
XX
XX 28-JUN-2001.
XX
XX 19-DEC-2000; 2000WO-US34584.
XX
XX 22-DEC-1999; 99US-0171785.
XX
XX (PARA-) PARADIGM GENETICS INC.
XX
XX Crawford JM, Rice J, Sevala V, Stewart S;
PI
XX
XX WPI; 2001-418081/44.
DR N-PSDB; AAH22802.
XX
XX Novel plant thioredoxin-porphobilinogen synthase or porphobilinogen
PT synthase polypeptides, useful for identifying compounds for use as
PT herbicides by inhibiting enzymatic activity of the polypeptides -
XX
XX Example 1; Page 19; 25pp; English.
XX
XX The invention provides novel DNA sequences encoding enzymes such as plant
CC thioredoxin-porphobilinogen synthase (T-PPS) and plant porphobilinogen
CC synthase (PPS). Methods of determining the enzymatic activity of T-PPS or
CC PPS or its functional fragment are provided that involves contacting
CC delta-aminolevulinic acid with the protein or its functional fragment and
CC measuring the amount of porphobilinogen formed from it. Compounds which
CC can modify the enzymatic activity T-PPS or PPS can also be identified
CC similarly, which are useful for inhibiting plant growth by inhibiting
CC enzymatic activity of T-PPS or PPS or its functional fragment. The
CC compounds thus identified are useful as herbicides. The present sequence
CC represents the thioredoxin functional fragment.
XX
XX
SQ Sequence 167 AA;
Query Match 11.6%; Score 50; DB 22; Length 167;
Best Local Similarity 100.0%; Pred. No. 6.3e-42;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDGDDDDKAMADIGS 50
DB 116 MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDGDDDDKAMADIGS 165

PA (ASTR) ASTRAZENECA AB.

PI Shaprio A, Sanyal G;

DR WPI; 2002-025884/03.

DR N-PSDB; AAH43782.

XX Production of a polypeptide delivery system useful as a medicament
PT comprises mixing together the polypeptide, cholesterol, saponin, and a
PT phospholipid with a polar head group, in the presence of a detergent

XX Disclosure; Page 45; 48pp; English.

XX The sequences given in ABA47728-32 are H. pylori HOP38 proteins
CC which were used in the method of the invention. The method of the
CC invention is a process for production of a polypeptide delivery system
CC comprising an immune stimulating complex (ISCOM) coupled to a
CC polypeptide of H. pylori or its antigenic fragment. The method comprises:
CC mixing the polypeptide, cholesterol, a saponin, and a phospholipid
CC having a polar head group; and removing the detergent from the mixture
CC to form an ISCOM. The method has a broad applicability to polypeptides,
CC including polypeptides that are unsuitable to prior art processes.

SQ Sequence 289 AA;

Query Match 11.6%; Score 50; DB 23; Length 289;
Best Local Similarity 100.0%; Pred. No. 1.1e-41;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRSGMKETAARKFERQHMDSPLGTDGDDKAMADIGS 50

|||||

DB 1 MHHHHHSSGLVPRSGMKETAARKFERQHMDSPLGTDGDDKAMADIGS 50

RESULT 12

ABG67266

ID ABG67266 standard; Protein; 402 AA.

XX AC ABG67266;

XX 24-SEP-2002 (first entry)

DE Polythoa 2 fluorescent protein encoded by plasmid pGR3.

XX Coral; Anthozoa species; fluorescence emission; fluorescence absorbance;

KW fluorescence resonance energy transfer; FRET; fluorescent labelling;

KW Polythoa 2 fluorescent protein; mutant; mutein.

XX Polythoa sp.

OS Synthetic.

XX WO200242323-A2.

XX 30-MAY-2002.

XX 22-NOV-2001; 2001WO-EPI3604.

XX 22-NOV-2000; 2000GB-0028495.

XX 22-NOV-2000; 2000US-252790P.

XX (DEVG-) DEVGEN NV.

XX Nys G, Plaetlinck G, Bogaert T;

XX WPI; 2002-500276/53.

XX N-PSDB; ABK96468.

XX Novel isolated fluorescent protein for producing fluorescence resonance
PT energy transfer for use in vivo labelling studies, is capable of
PT emitting fluorescence upon irradiation by incident light

XX Claim 7; Fig 4; 96pp; English.

XX

CC The present invention relates to novel coral fluorescent proteins
CC isolated from two brightly fluorescent Anthozoa species (Polythoa
CC and Discosoma species), and the polynucleotide sequences encoding
CC them. The fluorescent proteins of the invention are capable of
CC emitting fluorescence upon incident light irradiation, where the
CC maximal absorbance of the incident light is 440-480 nm, and maximal
CC fluorescence emission is 470-510 nm. The fluorescent proteins are
CC useful for producing fluorescence resonance energy transfer (FRET),
CC between a donor and acceptor fluorescent protein molecule. The
CC fluorescent proteins are useful for in vivo labelling studies, as
CC a label and/or marker, in particular as a genetic marker and/or
CC expression marker, in the fields of (micro-) biology, biochemistry
CC and/or molecular biology. They are also useful for monitoring
CC expression of proteins within biological systems and subcellular
CC localisation or trafficking of proteins. The polynucleotide
CC sequences encoding the fluorescent proteins are useful for in vitro
CC applications such as hybridisation and/or immunological assays. They
CC are also useful for labelling polypeptides and/or immunological assays. The
CC sequence represents a Polythoa 2 fluorescent protein encoded by a
CC plasmid as described in the examples of the present invention.

SQ Sequence 402 AA;

Query Match 11.6%; Score 50; DB 23; Length 402;

Best Local Similarity 100.0%; Pred. No. 1.4e-41;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRSGMKETAARKFERQHMDSPLGTDGDDKAMADIGS 50

|||||

DB 116 MHHHHHSSGLVPRSGMKETAARKFERQHMDSPLGTDGDDKAMADIGS 165

RESULT 13

ABG67267

ID ABG67267 standard; Protein; 407 AA.

XX AC ABG67267;

XX 24-SEP-2002 (first entry)

DE Polythoa 2 fluorescent protein encoded by plasmid pGR7.

XX Coral; Anthozoa species; fluorescence emission; fluorescence absorbance;

KW fluorescence resonance energy transfer; FRET; fluorescent labelling;

KW Polythoa 2 fluorescent protein; mutant; mutein.

XX Polythoa sp.

OS Synthetic.

XX WO200242323-A2.

XX 30-MAY-2002.

XX 22-NOV-2001; 2001WO-EPI3604.

XX 22-NOV-2000; 2000GB-0028495.

XX 22-NOV-2000; 2000US-252790P.

XX (DEVG-) DEVGEN NV.

XX Nys G, Plaetlinck G, Bogaert T;

XX WPI; 2002-500276/53.

XX N-PSDB; ABK96472.

XX Novel isolated fluorescent protein for producing fluorescence resonance
PT energy transfer for use in vivo labelling studies, is capable of
PT emitting fluorescence upon irradiation by incident light

XX Claim 7; Fig 5; 96pp; English.

XX The present invention relates to novel coral fluorescent proteins

CC isolated from two brightly fluorescent Anthozoa species (Polythoa

CC and *Discosoma* species), and the polynucleotide sequences encoding
 CC them. The fluorescent proteins of the invention are capable of
 CC emitting fluorescence upon incident light irradiation, where the
 CC maximal absorbance of the incident light is 440-480 nm, and maximal
 CC fluorescence emission is 470-510 nm. The fluorescent proteins are
 CC useful for producing fluorescence resonance energy transfer (FRET),
 CC between a donor and acceptor fluorescent protein molecule. The
 CC fluorescent proteins are useful for in vivo labelling studies, as
 CC a label and/or marker, in particular as a genetic marker and/or
 CC expression marker, in the fields of (micro-) biology, biochemistry
 CC and/or molecular biology. They are also useful for monitoring
 CC localisation or trafficking of proteins. The polynucleotide
 CC sequences encoding the fluorescent proteins are useful for in vitro
 CC applications such as hybridisation and/or immunological assays. They
 CC are also useful for labelling polypeptides of interest. The present
 CC sequence represents a *Polythoa* 2 fluorescent protein encoded by a
 CC plasmid as described in the examples of the present invention.
 XX
 SQ Sequence 407 AA;
 Query Match 11.6%; Score 50; DB 23; Length 407;
 Best Local Similarity 100.0%; Pred. No. 1.5e-41;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHHSSGLVPRGSGMKETAARFQHHMDSPDLGTDGDDDKAMADIGS 50
 Db 116 MHHHHHSSGLVPRGSGMKETAARFQHHMDSPDLGTDGDDDKAMADIGS 165
 RESULT 14
 ABG67270
 ID ABG67270 standard; Protein; 407 AA.
 XX
 AC ABG67270;
 XX
 DT 24-SEP-2002 (first entry)
 XX
 DE *Polythoa* 2 fluorescent protein encoded by plasmid pGR13.
 XX
 KW Coral; Anthozoa species; fluorescence emission; fluorescence absorbance;
 KW fluorescence resonance energy transfer; FRET; fluorescent labelling;
 KW *Polythoa* 2 fluorescent protein; mutant; mutagen.
 XX
 OS *Polythoa* sp.
 OS Synthetic.
 XX
 FN WO200242323-A2.
 XX
 PD 30-MAY-2002.
 XX
 PF 22-NOV-2001; 2001WO-EP13604.
 XX
 PR 22-NOV-2000; 2000GB-0028495.
 PR 22-NOV-2000; 2000US-252790P.
 XX
 PA (DEVG-) DEVGEN NV.
 XX
 PI Nys G, Plaetinck G, Bogaert T;
 XX
 DR WPI; 2002-500276/53.
 XX
 PT Novel isolated fluorescent protein for producing fluorescence resonance
 PT energy transfer for use in in vivo labelling studies, is capable of
 PT emitting fluorescence upon irradiation by incident light
 XX
 PS Claim 19; Fig 6; 96pp; English.
 XX
 CC The present invention relates to novel coral fluorescent proteins
 CC isolated from two brightly fluorescent Anthozoa species (*Polythoa*
 CC and *Discosoma* species), and the polynucleotide sequences encoding
 CC them. The fluorescent proteins of the invention are capable of
 CC emitting fluorescence upon incident light irradiation, where the

CC maximal absorbance of the incident light is 440-480 nm, and maximal
 CC fluorescence emission is 470-510 nm. The fluorescent proteins are
 CC useful for producing fluorescence resonance energy transfer (FRET),
 CC between a donor and acceptor fluorescent protein molecule. The
 CC fluorescent proteins are useful for in vivo labelling studies, as
 CC a label and/or marker, in particular as a genetic marker and/or
 CC expression marker, in the fields of (micro-) biology, biochemistry
 CC and/or molecular biology. They are also useful for monitoring
 CC localisation or trafficking of proteins. The polynucleotide
 CC sequences encoding the fluorescent proteins are useful for in vitro
 CC applications such as hybridisation and/or immunological assays. They
 CC are also useful for labelling polypeptides of interest. The present
 CC sequence represents a *Polythoa* 2 fluorescent protein encoded by a
 CC plasmid as described in the examples of the present invention.
 XX
 SQ Sequence 407 AA;
 Query Match 11.6%; Score 50; DB 23; Length 407;
 Best Local Similarity 100.0%; Pred. No. 1.5e-41;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHHSSGLVPRGSGMKETAARFQHHMDSPDLGTDGDDDKAMADIGS 50
 Db 116 MHHHHHSSGLVPRGSGMKETAARFQHHMDSPDLGTDGDDDKAMADIGS 165
 RESULT 15
 AAY85150
 ID AAY85150 standard; protein; 467 AA.
 XX
 AC AAY85150;
 XX
 DT 23-JUN-2000 (first entry)
 XX
 DE Mouse secreted protein acidic and rich in cysteine (SPARC).
 XX
 KW SPARC; secreted protein acidic and rich in cysteine; pharmaceutical;
 KW drug; research reagent; mouse.
 XX
 OS Mus musculus.
 XX
 PN JP3012931-B1.
 XX
 PD 28-FEB-2000.
 XX
 PF 26-FEB-1999; 99JP-0049826.
 XX
 PR 26-FEB-1999; 99JP-0049826.
 XX
 PA (AGEN) KOGYO GIJUTSUINCHO.
 XX
 DR WPI; 2000-259560/23.
 DR N-PSDB; AAA09770.
 XX
 PT New secretion protein acidic and rich in cysteine (SPARC) for drugs and
 PT research reagents consists of specific amino acid sequence -
 XX
 PS Claim 1; Fig 1; 17pp; Japanese.
 XX
 CC This sequence represents the amino acid sequence of a mouse protein,
 CC having the physiological activity of SPARC (secreted protein acidic and
 CC rich in cysteine). The protein is used for the preparation of drugs,
 CC research reagents and pharmaceuticals. The SPARC protein is prepared
 CC efficiently and inexpensively in large quantities. The protein can be
 CC used in pharmaceuticals as it does not lose its physiological activity.
 XX
 SQ Sequence 467 AA;
 Query Match 11.6%; Score 50; DB 21; Length 467;
 Best Local Similarity 100.0%; Pred. No. 1.7e-41;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRSGMKETAATAAFERQHMDSPDLGTDGDDKAMADIGS 50
|||||
Db 116 MHHHHHSSGLVPRSGMKETAATAAFERQHMDSPDLGTDGDDKAMADIGS 165

RESULT 16

AAAY79067
ID AAY79067 standard; Protein; 467 AA.

XX AC AAY79067;

XX DT 20-JUN-2000 (first entry)

XX DE Murine secreted protein acidic and rich in cysteine amino acid sequence.

XX KW Secreted protein acidic and rich in cysteine; SPARC; mouse; nerve cell;
XX KW neurocyte cell adhesion; cell protrusion retraction; cell shrinkage;
XX KW nervous system disease; epilepsy; arteriosclerosis; wound healing.

XX OS Mus sp.

XX PN JP3012930-B1.

XX PD 28-FEB-2000.

XX PF 26-FEB-1999; 99JP-0049708.

XX PR 26-FEB-1999; 99JP-0049708.

XX PA (AGEN) KOGYO GIUTSUINCHO.

XX DR WPI; 2000-306484/27.

XX DR N-PSDB; AAZ98759.

XX PT Drug composition for suppressing neurocyte cell adhesion, generating
XX PT cell migration and promoting shrinkage retraction of the cell contains
XX PT specific amino acid sequence.

XX PS Claim 1; Fig 4; 21pp; Japanese.

XX CC This sequence represents the amino acid sequence of the murine protein
XX CC referred to as SPARC (secreted protein acidic and rich in cysteine). The
XX CC protein is used as the active ingredient in a drug composition for
XX CC suppressing neurocyte cell adhesion, generating cell migration and
XX CC promoting a nerve protrusion shrinkage reaction. The composition can be
XX CC used in wound healing and also to treat diseases of the nervous system,
XX CC arteriosclerosis, and epilepsy.

SQ Sequence 467 AA;

Query Match 11.6%; Score 50; DB 21; Length 467;
Best Local Similarity 100.0%; Pred. No. 1.7e-41;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRSGMKETAATAAFERQHMDSPDLGTDGDDKAMADIGS 50
|||||
Db 116 MHHHHHSSGLVPRSGMKETAATAAFERQHMDSPDLGTDGDDKAMADIGS 165

RESULT 17

AAAB85251

ID AAB85251 standard; Protein; 551 AA.

XX AC AAB85251;

XX DT 07-SEP-2001 (first entry)

XX DE Plant thioedoxin-porphobilinogen synthase (T-PPS).

XX KW Thioedoxin; porphobilinogen synthase; T-PPS; pps; enzyme; herbicide;
XX KW delta-aminolevulinic acid; plant growth inhibitor.

XX OS Lycopersicon esculentum.

XX Key Location/Qualifiers
XX PH Protein 1..167
XX FT /note= "thioedoxin functional fragment"
XX FT 168..551
XX FT /note= "pps"

XX PN WO200146446-A1.

XX PD 28-JUN-2001.

XX PF 19-DEC-2000; 2000WO-US34584.

XX PR 22-DEC-1999; 99US-0171785.

XX PA (PARA-) PARADIGM GENETICS INC.

XX PI Crawford JM, Rice J, Sevala V, Stewart S;

XX DR WPI; 2001-418081/44.

XX DR N-PSDB; AAH22901.

XX PT Novel plant thioedoxin-porphobilinogen synthase or porphobilinogen
XX PT synthase polypeptides, useful for identifying compounds for use as
XX PT herbicides by inhibiting enzymatic activity of the polypeptides.

XX PS Disclosure; Page 16-17; 25pp; English.

XX CC The invention provides novel DNA sequences encoding enzymes such as plant
XX CC thioedoxin-porphobilinogen synthase (T-PPS) and plant porphobilinogen
XX CC synthase (PPS). Methods of determining the enzymatic activity of T-PPS or
XX CC PPS or its functional fragment are provided that involves contacting
XX CC delta-aminolevulinic acid with the protein or its functional fragment and
XX CC measuring the amount of porphobilinogen formed from it. Compounds which
XX CC can modify the enzymatic activity of T-PPS or PPS can also be identified
XX CC similarly, which are useful for inhibiting plant growth by inhibiting
XX CC enzymatic activity of T-PPS or PPS or its functional fragment. The
XX CC compounds thus identified are useful as herbicides. The present sequence
XX CC represents the fusion protein T-PPS.

SQ Sequence 551 AA;

Query Match 11.6%; Score 50; DB 22; Length 551;
Best Local Similarity 100.0%; Pred. No. 1.9e-41;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRSGMKETAATAAFERQHMDSPDLGTDGDDKAMADIGS 50
|||||

Db 116 MHHHHHSSGLVPRSGMKETAATAAFERQHMDSPDLGTDGDDKAMADIGS 165

RESULT 18

AAAB85252

ID AAB85252 standard; Protein; 551 AA.

XX AC AAB85252;

XX DT 07-SEP-2001 (first entry)

XX DE Plant thioedoxin-porphobilinogen synthase (T-PPS).

XX KW Thioedoxin; porphobilinogen synthase; T-PPS; pps; enzyme; herbicide;
XX KW delta-aminolevulinic acid; plant growth inhibitor.

XX OS Lycopersicon esculentum.

XX PH Protein 1..167

XX FT /note= "thioedoxin functional fragment"

XX FT 168..551

XX FT /note= "pps"

XX PN WO200146446-A1.

XX PD 28-JUN-2001.
 XX XX
 XX PF 19-DEC-2000; 2000WO-US34584.
 XX XX
 XX PR 22-DEC-1999; 99US-0171785.
 XX XX
 XX PA (PARA-) PARADIGM GENETICS INC.
 XX XX
 XX PI Crawford JM, Rice J, Sevala V, Stewart S;
 XX XX
 XX DR WPI; 2001-418081/44.
 XX DR N-PSDB; AAH22801.
 XX XX
 XX PT Novel plant thioredoxin-porphobilinogen synthase or porphobilinogen
 XX PT synthase polypeptides, useful for identifying compounds for use as
 XX PT herbicides by inhibiting enzymatic activity of the polypeptides -
 XX XX
 XX PS Claim 4; Page 17-18; 25pp; English.
 XX XX
 XX CC The invention provides novel DNA sequences encoding enzymes such as plant
 XX CC thioredoxin-porphobilinogen synthase (T-PPS) and plant porphobilinogen
 XX CC synthase (PPS). Methods of determining the enzymatic activity of T-PPS or
 XX CC PPS or its functional fragment are provided that involve contacting
 XX CC delta-aminolevulinic acid with the protein or its functional fragment and
 XX CC measuring the amount of porphobilinogen formed from it. Compounds which
 XX CC can modify the enzymatic activity of T-PPS or PPS can also be identified
 XX CC similarly, which are useful for inhibiting plant growth by inhibiting
 XX CC enzymatic activity of T-PPS or PPS or its functional fragment. The
 XX CC compounds thus identified are useful as herbicides. The present sequence
 XX CC represents the fusion protein T-PPS.
 XX XX
 XX SQ Sequence 551 AA;
 Query Match 11.6%; Score 50; DB 22; Length 551;
 Best Local Similarity 100.0%; Pred. No. 1.9e-41;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHHSSGLVPRGSGMKETAARFQHMDSPLGTDKAMADIGS 50
 DB 116 MHHHHHSSGLVPRGSGMKETAARFQHMDSPLGTDKAMADIGS 165
 RESULT 19
 ID AAB11532 standard; Protein; 692 AA.
 XX AAB11532;
 XX AC AAB11532;
 XX DT 12-OCT-2000 (first entry)
 XX DE SEN virus protein fragment SEQ ID NO: 104.
 XX KW SEN virus; SENV; gastrointestinal tract disorder; inflammatory disease;
 XX KW proliferative disorder; hepatopathy; hepatitis; viral infection;
 XX KW vaccination; gene therapy.
 XX OS Hepatitis virus.
 XX PN WO200028039-A2.
 XX PD 18-MAY-2000.
 XX XX
 XX PF 09-NOV-1999; .99WO-EF08566.
 XX XX
 XX PR 10-NOV-1998; 98IT-MI02437.
 XX PR 30-APR-1999; 99IT-MI00923.
 XX PR 14-MAY-1999; 99EP-0830298.
 XX PR 16-JUL-1999; 99EP-0113932.
 XX XX
 XX PA (DIAS-) DIASORIN SRL.
 XX XX
 XX PI Primi D, Fiordalisi G, Mantero GL, Mattioli S, Sottini A;

PI Bonelli F, Vaglini L, Olivero P, Dal Corso A, Bonelli M;
 XX DR WPI; 2000-376551/32.
 XX XX
 XX PT Nucleic acids representing the genome of the SEN virus (SENV) and
 XX PT encoded proteins, useful for treatment of hepatopathies, inflammatory
 XX PT diseases and proliferative disorders such as cancer -
 XX XX
 XX PS Example 16; Page 78; 392pp; English.
 XX XX
 XX CC The present invention is concerned with the sequence of the genome of the
 XX CC SEN virus (SENV), and the proteins encoded by it. SENV is thought to be
 XX CC the cause of hepatopathies which are not linked to the presence of the
 XX CC hepatitis A, B and E viruses in man. The genome and proteins of this
 XX CC virus can be used in gene therapy and vaccination against the virus,
 XX CC which also causes disorders of the gastrointestinal tract, including
 XX CC Crohn's disease and lupus erythematosus, inflammatory diseases, and
 XX CC proliferative disorders such as cancer.
 XX XX
 XX SQ Sequence 692 AA;
 Query Match 11.6%; Score 50; DB 21; Length 692;
 Best Local Similarity 100.0%; Pred. No. 2.4e-41;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHHSSGLVPRGSGMKETAARFQHMDSPLGTDKAMADIGS 50
 DB 1 MHHHHHSSGLVPRGSGMKETAARFQHMDSPLGTDKAMADIGS 50
 RESULT 20
 ID AAM48247 standard; Protein; 824 AA.
 XX AAM48247;
 XX AC AAM48247;
 XX DT 21-MAR-2002 (first entry)
 XX DE Thioredoxin/deoxyxylulose 5-phosphate synthase fusion protein.
 XX KW DXPS: deoxyxylulose 5-phosphate synthase; pyruvate; thioredoxin;
 XX KW glyceraldehyde 3-phosphate; plant growth modulator;
 XX KW microbial growth modulator; enzyme.
 XX OS Chimeric - Arabidopsis thaliana.
 XX OS Chimeric - Synthetic.
 XX FH Key Location/Qualifiers
 XX FT Region 1..165
 XX FT /note= "Thioredoxin protein"
 XX FT Region 166..824
 XX FT /note= "Truncated DXPS protein"
 XX XX
 XX PN US6326164-B1.
 XX XX
 XX PD 04-DEC-2001.
 XX XX
 XX PF 27-JUL-2000; 2000US-0626589.
 XX XX
 XX PR 27-JUL-2000; 2000US-0626589.
 XX XX
 XX PA (PARA-) PARADIGM GENETICS INC.
 XX XX
 XX PI Rice JW, Kloti AS, Crawford JM, Lanning B, Stewart SJ;
 XX XX
 XX DR WPI; 2002-121106/16.
 XX DR N-PSDB; ABA95618.
 XX XX
 XX PT Assay for determining deoxyxylulose 5-phosphate synthase activity by
 XX PT measuring pyruvate depletion, useful for screening compounds that
 XX PT inhibit or enhance this activity which is useful for modulating plant
 XX PT and microbial growth -
 XX XX

PS Claim 1; Columns 19-24; 26pp; English.

XX The present sequence is a fusion protein comprising a truncated

CC deoxyxylulose 5-phosphate synthase (DXPS) from Arabidopsis thaliana and

CC thioredoxin (trxA). The N-terminal 58 amino acids of the DXPS protein

CC were removed to generate the truncated protein. The truncated DXPS protein

CC possesses DXPS activity. The full-length DXPS is given in AAM48245. The

CC present invention relates to a method for determining DXPS activity. The

CC method comprises contacting pyruvate, and optionally glyceraldehyde

CC 3-phosphate with DXPS and then determining the concentration of remaining

CC pyruvate and/or glyceraldehyde 3-phosphate. The method is useful for

CC screening for inhibitors and enhancers of DXPS activity which will have

XX use as modulators of plant and microbial growth.

SQ Sequence 824 AA;

Query Match 11.6%; Score 50; DB 23; Length 824;

Best Local Similarity 100.0%; Pred. No. 2.8e-41;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHSSGLVPRGSGMKETAARFQRHMDSPDLGTDGDDDDKAMADIGS 50

Db 116 MHHHHHSSGLVPRGSGMKETAARFQRHMDSPDLGTDGDDDDKAMADIGS 165

RESULT 21

AAR41355

ID AAR41355 standard; peptide; 49 AA.

XX

AC AAR41355;

XX

DT 04-MAR-1994 (first entry)

XX

DE MSP1EGF1B EGF1-like domain variant.

XX

KW Epidermal growth factor 1; merozoite surface protein 1; malaria;

KW vaccine.

XX

OS Plasmodium yoellii.

XX

FH Key Location/Qualifiers

FT Cleavage-site 1

FT /note= "introduced to facilitate cleavage

FT from recombinant protein"

XX

PN W09317107-A.

XX

XX 02-SEP-1993.

XX

PF 22-FEB-1993; 93WO-GB00367.

XX

PR 22-FEB-1992; 92GB-0003821.

XX

PA (MEDI-) MEDICAL RES COUNCIL.

XX

PI Blackman MJ, Chappel JA, Holder AA;

XX

DR WPI; 1993-288413/36.

XX

PT Allelic variants of epidermal growth factor 1- or 2-like domains - of

PT merozoite surface protein 1, produced recombinantly for malaria

PT vaccines

XX

PS Claim 1; Fig 1b; 35pp; English.

XX The sequence is that of an allelic variant of a merozoite surface

CC protein-1 epidermal growth factor (EGF) 1 like domain. It may be

CC used alone or as part of a fusion protein of EGF-1-like and

CC EGF-2-like domains in vaccines against malaria. When expressed

CC recombinantly it is produced in a form indistinguishable from

CC that in the native protein.

XX

SQ Sequence 49 AA;

Query Match 11.1%; Score 48; DB 14; Length 49;

Best Local Similarity 100.0%; Pred. No. 2e-40;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 338 NISQHCQVKKQCPENSGCFRHLDERECKCLLNYYKQEGDKCVENPNT 385

Db 2 NISQHCQVKKQCPENSGCFRHLDERECKCLLNYYKQEGDKCVENPNT 49

RESULT 22

AAB37608

ID AAB37608 standard; protein; 96 AA.

XX

AC AAB37608;

XX

DT 27-FEB-2001 (first entry)

XX

DE Merozoite surface protein-1.

XX

KW Merozoite surface protein; protozoacide; vaccine; malaria.

XX

OS Plasmodium falciparum.

XX

PN W0200063245-A2.

XX

PD 26-OCT-2000.

XX

PF 20-APR-2000; 2000WO-GB01558.

XX

PR 20-APR-1999; 99GB-0009072.

PR 13-MAY-1999; 99US-0311817.

PR 25-MAY-1999; 99CA-2271451.

XX

PA (MEDI-) MEDICAL RES COUNCIL.

XX

PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaiipibull C;

XX

DR WPI; 2001-015762/02.

XX

PT Novel variants of the C-terminal fragment of Plasmodium merozoite

PT surface protein-1, useful as vaccines for treating or preventing

PT malaria

XX

PS Example 2; Page 48; 126pp; English.

XX The present invention relates to non-natural variants of a C-terminal

CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The

CC non-natural variants have reduced affinity for at least 1 antibody

CC capable of blocking a second antibody that inhibits the proteolytic

CC cleavage of Plasmodium MSP-1_4_2, and has the same affinity for at least

CC one third antibody that inhibits the proteolytic cleavage of Plasmodium

CC MSP-1_4_2, compared to natural MSP-1_1_9. The present sequence is the

CC wild-type MSP-1 protein. This sequence was used to generate the variants

CC of the present invention. The non-natural variants of the present

CC invention are useful for immunising a mammal against malaria, and can be

CC used to treat malaria.

XX

SQ Sequence 96 AA;

Query Match 10.7%; Score 46; DB 22; Length 96;

Best Local Similarity 100.0%; Pred. No. 4e-38;

Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 NSGCFRHLDERECKCLLNYYKQEGDKCVENPNTCNENNGGCDADA 397

Db 15 NSGCFRHLDERECKCLLNYYKQEGDKCVENPNTCNENNGGCDADA 60

RESULT 23

AAB37609

ID AAB37609 standard; Protein; 108 AA.

XX

```
AC AAB37609;
XX
XX 27-FEB-2001 (first entry)
XX
XX Merozoite surface protein-119.
XX
XX Merozoite surface protein; protozoacide; vaccine; malaria.
XX
XX Plasmodium falciparum.
XX
XX WO200063245-A2.
XX
XX 26-OCT-2000.
XX
XX 20-APR-2000; 2000WO-GB01558.
XX
XX 20-APR-1999; 99GB-0009072.
XX
XX 13-MAY-1999; 99US-0311817.
XX
XX 23-MAY-1999; 99CA-2271451.
XX
XX (MEDI-) MEDICAL RES COUNCIL.
XX
XX Holder A, Birdsell B, Feeney J, Morgan W, Syed S, Uthaipibull C;
XX
XX WPI; 2001-015762/02.
XX
XX N-PSDB; AAC68977.
XX
XX Novel variants of the C-terminal fragment of Plasmodium merozoite
XX surface protein-1, useful as vaccines for treating or preventing
XX malaria.
XX
XX Example 5; Fig 15; 126pp; English.
XX
XX The present invention relates to non-natural variants of a C-terminal
XX fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
XX non-natural variants have reduced affinity for at least 1 antibody
XX capable of blocking a second antibody that inhibits the proteolytic
XX cleavage of Plasmodium MSP-14-2, and has the same affinity for at least
XX one third antibody that inhibits the proteolytic cleavage of Plasmodium
XX MSP-14-2, compared to natural MSP-11-9. The non-natural variants of the
XX present invention are useful for immunising a mammal against malaria, and
XX can be used to treat malaria. The present sequence is MSP-119 protein.
XX
XX Sequence 108 AA;
XX
XX Query Match 10.7%; Score 46; DB 22; Length 108;
XX Best Local Similarity 100.0%; Pred. No. 4.5e-38;
XX Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 352 NSGCFRHLDERECKLLNYKQEGDKCVENPNTCNENGGCDADA 397
XX |||||||||||||||||||||||||||||||||||||||||||||||
XX DB 27 NSGCFRHLDERECKLLNYKQEGDKCVENPNTCNENGGCDADA 72
XX
XX RESULT 24
XX AAB11534
XX ID AAB11534 standard; Protein; 138 AA.
XX
XX AAB11534;
XX
XX 12-OCT-2000 (first entry)
XX
XX SEN virus protein fragment SEQ ID NO: 108.
XX
XX SEN virus; SEN; gastrointestinal tract disorder; inflammatory disease;
XX proliferative disorder; hepatopathy; hepatitis; viral infection;
XX vaccination; gene therapy.
XX
XX Hepatitis virus.
XX
XX WO200028039-A2.
XX
XX 18-MAY-2000.
XX
XX PD
```

```
XX 09-NOV-1999; 99WO-EF08566.
XX
XX 10-NOV-1998; 98IT-MI02437.
XX
XX 30-APR-1999; 99IT-MI00923.
XX
XX 14-MAY-1999; 99EP-0830298.
XX
XX 16-JUL-1999; 99EP-0113932.
XX
XX (DIAS-) DIASORIN SRL.
XX
XX Primi D, Fiordalisi G, Mantero GL, Mattioli S, Sottini A;
XX Bonelli F, Vaglini L, Olivero P, Dal Corso A, Bonelli M;
XX WPI; 2000-376551/32.
XX
XX Nucleic acids representing the genome of the SEN virus (SENV) and
XX encoded proteins, useful for treatment of hepatopathies, inflammatory
XX diseases and proliferative disorders such as cancer.
XX
XX Example 16; Page 80; 392pp; English.
XX
XX The present invention is concerned with the sequence of the genome of the
XX SEN virus (SENV), and the proteins encoded by it. SENV is thought to be
XX the cause of hepatopathies which are not linked to the presence of the
XX hepatitis A, B and E viruses in man. The genome and proteins of this
XX virus can be used in gene therapy and vaccination against the virus,
XX which also causes disorders of the gastrointestinal tract, including
XX Crohn's disease and lupus erythematosus, inflammatory diseases, and
XX proliferative disorders such as cancer.
XX
XX Sequence 138 AA;
XX
XX Query Match 10.7%; Score 46; DB 21; Length 138;
XX Best Local Similarity 100.0%; Pred. No. 5.6e-38;
XX Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MHHHHHSSGLVPRGSGMKETAARQHMDSPLDGTDDDDKAMA 46
XX |||||||||||||||||||||||||||||||||||||||||||||||
XX DB 1 MHHHHHSSGLVPRGSGMKETAARQHMDSPLDGTDDDDKAMA 46
XX
XX RESULT 25
XX AAE20460
XX ID AAE20460 standard; Protein; 284 AA.
XX
XX AAE20460;
XX
XX 01-JUL-2002 (first entry)
XX
XX Recombinant (SPKR)4Inv protein.
XX
XX Gene delivery vehicle; nucleic acid binding domain; (SPKR)4Inv protein;
XX integrin binding domain; modular protein; transgenic; integrin.
XX
XX Unidentified.
XX
XX Key Location/Qualifiers
XX Region 2..7 /note= "His tag"
XX
XX Cleavage-site 11..16 /note= "Thrombin cleavage site"
XX
XX Region 19..33 /note= "S tag"
XX
XX Cleavage-site 39..43 /note= "Enterokinase cleavage site"
XX
XX Domain 47..62 /note= "DNA binding domain"
XX
XX Region 66..86 /note= "Linker"
XX
XX Region 89..284 /note= "Invasin"
XX
XX EP1178117-A1.
XX
XX PD
```

```
XX PD 06-FEB-2002.
XX PF 02-AUG-2000; 2000EP-0202750.
XX PR 02-AUG-2000; 2000EP-0202750.
XX PA (UYRO-) UNIV ROTTERDAM ERASMUS.
XX PI Scholte BJ;
XX DR WPI: 2002-270869/32.
XX DR N-PSDB; AAD32715.
XX PT New gene delivery vehicle comprising a nucleic acid binding domain and
XX PT an integrin binding domain, useful for delivering therapeutic molecule
XX PT to a cell exposing integrin and enhance efficacy of gene delivery using
XX PT integrin receptors
XX PS Disclosure; Fig 1A; 32pp; English.
XX CC The invention relates to a gene delivery vehicle comprising a general
XX CC nucleic acid binding domain and an integrin binding domain. The
XX CC invention also provides modular protein (SPKR)4inv, designed to mediate
XX CC integrin targeted gene delivery. The gene delivery vehicle is useful
XX CC for delivering therapeutic molecule to a cell exposing integrin, and
XX CC subsequently enhance the efficacy of gene delivery using integrin
XX CC receptors. The present sequence is recombinant (SPKR)4inv protein.
XX SQ Sequence 284 AA;
XX Query Match 10.7%; Score 46; DB 23; Length 284;
XX Best Local Similarity 100.0%; Pred. No. 1.1e-37;
XX Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHSSGLVPRGSGMKETAARFERQHMDSPDLGTDGDDKAMA 46
DB 1 MHHHHHSSGLVPRGSGMKETAARFERQHMDSPDLGTDGDDKAMA 46
RESULT 26
AAY09372
ID AAY09372 standard; Protein; 355 AA.
XX AC AAY09372;
XX DT 31-AUG-1999 (first entry)
XX DE Merozoite surface protein MSP-1-42.
XX KW MSP-1; merozoite surface protein; malaria; vaccine;
XX KW protein engineering; protein expression; codon usage;
XX KW transgenic animal.
XX OS Plasmodium falciparum.
XX PN WO9920774-A2.
XX PD 29-APR-1999.
XX PF 20-OCT-1998; 98WO-US22226.
XX PR 15-MAY-1998; 98US-0085649.
XX PR 20-OCT-1997; 97US-0062592.
XX PA (GENZ ) GENZYME TRANSGENICS CORP.
XX PI Chen LH, Meade H;
XX DR WPI: 1999-288313/24.
XX DR P-PSDB; AAX56008.
XX PT Modified malarial protein for use in anti-malarial vaccines
```

```
XX PS Example; Fig 1; 35pp; English.
XX CC The present sequence represents a 42 kDa C-terminal portion of
XX CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
XX CC important target for the development of a vaccine against
XX CC Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42
XX CC has been modified (see AAX56008) compared to the native sequence (see
XX CC AAX56009) such that 306 nucleotide positions have been replaced to
XX CC lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
XX CC instability motifs while maintaining the same protein amino acid
XX CC sequence. These alterations allow MSP-1-42 to be expressed in
XX CC mammalian cell culture and in transgenic mice. Native MSP-1-12
XX CC is known to be difficult to express in cell culture systems,
XX CC mammalian cell culture systems or in transgenic animals. The
XX CC invention allows expression of MSP-1 protein in the milk of
XX CC transgenic animals, and also provides a DNA vaccine comprising a
XX CC vector containing the altered MSP-1-42 sequence.
XX SQ Sequence 355 AA;
XX Query Match 10.7%; Score 46; DB 20; Length 355;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-37;
XX Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 352 NSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGGCDADA 397
DB 277 NSGCFRHLDEREECKLLNYKQEGDKCVENPNPTCNENNGGCDADA 322
RESULT 27
AAY05832
ID AAY05832 standard; Protein; 355 AA.
XX AC AAY05832;
XX DT 02-AUG-1999 (first entry)
XX DE Merozoite surface protein MSP-1-42.
XX KW MSP-1; merozoite surface protein; malaria; vaccine;
XX KW protein engineering; protein expression; codon usage;
XX KW transgenic animal.
XX OS Plasmodium falciparum.
XX PN WO9920766-A2.
XX PD 29-APR-1999.
XX PF 20-OCT-1998; 98WO-US22225.
XX PR 15-MAY-1998; 98US-0085649.
XX PR 20-OCT-1997; 97US-0062592.
XX PA (GENZ ) GENZYME TRANSGENICS CORP.
XX PI Chen LH, Meade H;
XX DR WPI: 1999-302742/25.
XX DR N-PSDB; AAX25586.
XX PT New modified recombinant nucleic acid sequences useful for producing
XX PT malarial DNA vaccine
XX PS Disclosure; Fig 1; 43pp; English.
XX CC The present sequence represents a 42 kDa C-terminal portion of
XX CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
XX CC important target for the development of a vaccine against
XX CC Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42
XX CC has been modified (see AAX25586) compared to the native sequence (see
XX CC AAX25587) such that 306 nucleotide positions have been replaced to
```

CC lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
 CC instability motifs while maintaining the same protein amino acid
 CC sequence. These alterations allow MSP-1-42 to be expressed in
 CC mammalian cell culture and in transgenic mice. The invention
 CC provides modified recombinant nucleic acid sequences and methods for
 CC increasing the mRNA levels and protein expression of proteins that
 CC are difficult to express in cell culture systems, mammalian cell
 CC culture systems or in transgenic animals. The preferred difficult
 CC protein candidates for expression are those derived from lower
 CC organisms such as parasites, bacteria and viruses that have DNA
 CC coding sequences of high AT content or which have mRNA instability
 CC motifs or rare codons relative to the recombinant expression system
 CC to be used. The invention allows expression of MSP-1 protein in
 CC the milk of transgenic animals, and also provides a DNA vaccine
 CC comprising a vector containing the altered MSP-1-42 sequence.

XX Sequence 355 AA;

Query Match 10.7%; Score 46; DB 20; Length 355;
 Best Local Similarity 100.0%; Pred. No. 1.4e-37;
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 NSGCFRHLDERECKLLNYKQEGDKCVENPPTCENNNGCCDADA 397

DB 277 NSGCFRHLDERECKLLNYKQEGDKCVENPPTCENNNGCCDADA 322

RESULT 28

AAV09373

ID AAY09373 standard; Protein; 361 AA.

AC AAY09373;

XX 31-AUG-1999 (first entry)

DE Merozoite surface protein MSP-1-42.

XX MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 KW transgenic animal.

OS Plasmodium falciparum.

XX WO9920774-A2.

XX 29-APR-1999.

XX 20-OCT-1998; 98WO-US222226.

XX 15-MAY-1998; 98US-0085649.

XX 20-OCT-1997; 97US-0062592.

XX (GENZ) GENZYME TRANSGENICS CORP.

XX Chen LH, Meade H;

XX WPI; 1999-288313/24.

XX P-PSDB; AAX56009.

XX Modified malarial protein for use in anti-malarial vaccines

XX Example; Fig 2; 35pp; English.

XX This present sequence comprises a 42 kDa C-terminal portion of
 CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
 CC important target for the development of a vaccine against
 CC Plasmodium falciparum. The C-terminal end of the sequence is
 CC modified to include a 6xHis tag. A nucleic acid (see AAX56009)

CC encoding MSP-1-42 has been modified according to a method
 CC of the invention in order to improve expression in mammalian cells
 CC and in transgenic animals by reducing the AT content and removing
 CC mRNA instability motifs. The invention allows expression of
 CC MSP-1-42 in the milk of transgenic animals, and also provides a DNA

CC vaccine comprising a vector containing the altered MSP-1-42 nucleic
 CC acid.
 XX Sequence 361 AA;

Query Match 10.7%; Score 46; DB 20; Length 361;
 Best Local Similarity 100.0%; Pred. No. 1.4e-37;
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 NSGCFRHLDERECKLLNYKQEGDKCVENPPTCENNNGCCDADA 397

DB 277 NSGCFRHLDERECKLLNYKQEGDKCVENPPTCENNNGCCDADA 322

RESULT 29

AAV05833

ID AAY05833 standard; Protein; 361 AA.

XX AAY05833;

XX 02-AUG-1999 (first entry)

XX Merozoite surface protein MSP-1-42.

XX MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 KW transgenic animal.

XX Plasmodium falciparum.

XX WO9920766-A2.

XX 29-APR-1999.

XX 20-OCT-1998; 98WO-US222225.

XX 15-MAY-1998; 98US-0085649.

XX 20-OCT-1997; 97US-0062592.

XX (GENZ) GENZYME TRANSGENICS CORP.

XX Chen LH, Meade H;

XX WPI; 1999-302742/25.

XX N-PSDB; AAX25587.

XX New modified recombinant nucleic acid sequences useful for producing
 PT malarial DNA vaccine

XX Disclosure; Fig 2; 43pp; English.

XX This present sequence comprises a 42 kDa C-terminal portion of
 CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
 CC important target for the development of a vaccine against
 CC Plasmodium falciparum. The C-terminal end of the sequence is
 CC modified to include a 6xHis tag. Nucleic acids (see AAX25586 and
 CC AAX25593) encoding MSP-1-42 have been modified according to a method
 CC of the invention in order to improve expression in mammalian cells
 CC and in transgenic animals. The invention provides modified
 CC recombinant nucleic acid sequences and methods for increasing the
 CC mRNA levels and protein expression of proteins that are difficult
 CC to express in cell culture systems, especially mammalian cell
 CC culture systems or in transgenic animals. The preferred difficult
 CC protein candidates for expression are those derived from lower
 CC organisms such as parasites, bacteria and viruses that have DNA
 CC coding sequences of high AT content or which have mRNA instability
 CC motifs or rare codons relative to the recombinant expression system
 CC to be used. The invention allows expression of MSP-1 in the milk
 CC of transgenic animals, and also provides a DNA vaccine comprising a
 CC vector containing the altered MSP-1-42 nucleic acid.

XX Sequence 361 AA;

Qy 352 NSGCFRHLDEREECKCLLNKQEGDKCVENPNPTCENNNGGCDADA 397

Query Match 10.7%; Score 46; DB 20; Length 376;
Best Local Similarity 100.0%; Pred. No. 1.5e-37;

Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 NSGCFRHLDERECKCLLYKQEGDKCVENPNTCNENNGGCDADA 397
 |||||
 Db 292 NSGCFRHLDERECKCLLYKQEGDKCVENPNTCNENNGGCDADA 337

RESULT 32
 AAW84315
 ID AAW84315 standard; Protein; 408 AA.
 AC AAW84315;
 XX
 DT 22-MAR-1999 (first entry)
 XX
 DE TrxA-rabbit tissue factor fusion protein.
 XX
 KW Thiorodoxin A protein; rabbit tissue factor; RTF; trxA;
 KW blood coagulation cascade; fusion protein; chimeric;
 KW standard blood clotting test; blood disorder; blood coagulation;
 KW prothrombin time test.
 XX
 OS Chimeric - Oryctolagus cuniculus.
 OS Chimeric - Escherichia coli.
 OS Chimeric - Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Protein 1..109
 FT /note= "Thiorodoxin A protein"
 FT Cleavage-site 126..131
 FT /note= "thrombin cleavage site"
 FT Peptide 134..148
 FT /note= "S-peptide marker"
 FT Cleavage-site 154..158
 FT /note= "enterkinase cleavage site"
 FT Protein 169..408
 FT /note= "truncated rabbit tissue factor"
 XX
 PN US5858724-A.
 XX
 PD 12-JAN-1999.
 XX
 PF 16-JUL-1996; 96US-0683007.
 XX
 PR 16-JUL-1996; 96US-0683007.
 XX
 PA (PELF-) PEL FREEZ.
 XX
 PI Dománico MJ, Kroecker W, Novy RE, Yaeger KW;
 XX WPI; 1999-119880/10.
 DR N-PSDB; AAV63788.
 XX
 PT Fusion protein containing; truncated rabbit tissue factor;
 PT thiorodoxin A protein and; a tag for affinity purification - useful
 PT in blood clotting assays and can be produced as a soluble protein in
 PT the correctly folded form
 XX
 PS Claim 11; Columns 13-16; 13pp; English.
 XX
 CC The present sequence represents a fusion protein that comprises, in
 CC an N to C terminal direction, a thiorodoxin A (trxA) protein from
 CC Escherichia coli, a His marker, a S-peptide and a truncated rabbit
 CC tissue factor (RTF) sequence that includes the extracellular and
 CC transmembrane domains, but not the cytoplasmic domain, of a native
 CC mature RTF. RTF can function as part of a blood coagulation cascade.
 CC The fusion protein is useful in standard blood clotting tests performed
 CC before surgery to identify disorders of blood coagulation, e.g. the
 CC prothrombin time test.
 XX
 SQ Sequence 408 AA;

Query Match 10.7%; Score 46; DB 20; Length 408;

Best Local Similarity 100.0%; Pred. No. 1.6e-37;
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHSSGLVPRGSGMKETAATAKFERQHRMDSPLDGTDDDDKAWA 46
 |||||
 Db 116 MHHHHHSSGLVPRGSGMKETAATAKFERQHRMDSPLDGTDDDDKAWA 161

RESULT 33
 AAW54145
 ID AAW54145 standard; Protein; 1639 AA.
 AC AAW54145;
 XX
 DT 23-SEP-1998 (first entry)
 XX
 DE P. falciparum synthetic gpl90 protein.
 XX
 KW gpl90; malaria; MSP-1; merozoite surface protein; stability; vaccine;
 KW monoclonal antibody; passive immunisation; parasite.
 XX
 OS Plasmodium falciparum.
 OS Synthetic.
 XX
 PN WO9814583-A2.
 XX
 PD 09-APR-1998.
 XX
 PF 02-OCT-1997; 97WO-EP05441.
 XX
 PR 02-OCT-1996; 96DE-4040817.
 XX
 PA (BUJA/) BUJARD H.
 XX
 PI Bujard H, Pan W, Tolle R;
 XX WPI; 1998-240088/21.
 DR N-PSDB; AAV21451, AAV35363.
 XX
 PT Recombinant production of complete gp190/MSP-1 Plasmodium surface
 PT protein - useful in anti-malaria vaccines, also stabilising genes by
 PT -reducing their AT content
 XX
 PS Example 1; Fig 3c; 48pp; German.
 XX
 CC This sequence represents a modified Plasmodium falciparum gp190/MSP-1
 CC (merozoite surface) protein. The gene encoding this protein has been
 CC stabilised by reducing the AT content of the nucleotide sequence. Such a
 CC protein is useful in vaccines against malaria or for producing monoclonal
 CC antibodies (for passive immunisation). The complete gp190 protein can now
 CC be produced outside the parasite and has, at least over extended regions,
 CC the native pattern of folding. Larger amounts of the protein can be
 CC produced recombinantly than would be possible using the parasites as
 CC source.
 XX
 SQ Sequence 1639 AA;

Query Match 10.7%; Score 46; DB 19; Length 1639;
 Best Local Similarity 100.0%; Pred. No. 5.9e-37;
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 NSGCFRHLDERECKCLLYKQEGDKCVENPNTCNENNGGCDADA 397
 |||||
 Db 1540 NSGCFRHLDERECKCLLYKQEGDKCVENPNTCNENNGGCDADA 1585

RESULT 34
 ABB80661
 ID ABB80661 standard; peptide; 59 AA.
 XX
 AC ABB80661;
 XX
 DT 15-JUL-2002 (first entry)

PN WO9317107-A.
 XX 02-SEP-1993.
 XX 22-FEB-1993; 93WO-GB00367.
 XX 22-FEB-1992; 92GB-0003821.
 XX (MEDI-) MEDICAL RES COUNCIL.
 XX Blackman MJ, Chappel JA, Holder AA;
 PI WPT; 1993-288413/36.
 DR
 XX
 PT Allelic variants of epidermal growth factor 1- or 2-like domains - of
 PT merozoite surface protein 1, produced recombinantly for malaria
 PT vaccines
 XX
 PS Claim 1; Fig 1a; 35pp; English.
 CC The sequence is that of an allelic variant of a merozoite surface
 CC protein-1 epidermal growth factor (EGF) 1 like domain. It may be
 CC used alone or as part of a fusion protein of EGF-1-like and
 CC EGF-2-like domains in vaccines against malaria. When expressed
 CC recombinantly it is produced in a form indistinguishable from
 CC that in the native protein.
 XX
 SQ Sequence 49 AA;
 Query Match 7.9%; Score 34; DB 14; Length 49;
 Best Local Similarity 100.0%; Pred. No. 2.7e-26;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 352 NSGCFRHLDERECKCLNLYKQEGDKCVENPNT 385
 DB 16 NSGCFRHLDERECKCLNLYKQEGDKCVENPNT 49
 RESULT 37
 AAB51128
 ID AAB51128 standard; Protein: 55 AA.
 AC AAB51128;
 XX
 DT 20-MAR-2001 (first entry)
 XX
 DE Human mamaglobin N-terminal fragment SEQ ID NO:28.
 XX
 KW Human; mamaglobin; breast cancer; detection; diagnosis; antibody;
 KW vaccine; cytostatic; antinamaglobin.
 XX
 OS Homo sapiens.
 XX
 PN WO200073338-A1.
 XX
 PD 07-DEC-2000.
 XX
 PF 26-MAY-2000; 2000WO-US14845.
 XX
 PR 28-MAY-1999; 99US-0136528.
 PR 01-JUN-1999; 99US-0137048.
 XX
 PA (CORI-) CORIXA CORP.
 PA (HEND/) HENDRICKSON R C.
 PA (HOUG/) HOUGHTON R L.
 PA (REED/) REED S G.
 XX
 PI Fanger GR;
 XX
 DR WPT; 2001-049928/06.
 XX
 PT Polypeptide comprising at least seven consecutive amino acid residues
 PT of human mamaglobin, useful in the treatment and detection of breast

PT cancer -
 XX
 PS
 XX Example 1; Fig 2; 109pp; English.
 CC The present invention describes human mamaglobin peptides (I)
 CC comprising at least 7 consecutive residues. Also described are: (1) a
 CC vaccine comprising (I) with an immunostimulant which is an adjuvant;
 CC (2) an isolated antibody (Ab1) or its antigen-binding fragment, which
 CC specifically binds to a mamaglobin epitope having the sequence of
 CC pro2-3; (3) an isolated antibody (Ab2) or its antigen-binding fragment
 CC that specifically binds to glycosylated mamaglobin; (4) a method for
 CC inhibiting the development of breast cancer in a patient, comprising
 CC administering (I) or Ab1 or Ab2; and (5) a method (M1) for determining
 CC the presence or absence of breast cancer in a patient. (I) has
 CC cytotstatic activity. The polypeptides and antibodies are used in
 CC vaccines for the prevention and treatment of breast cancer. They are
 CC also used for diagnosis and monitoring of breast cancer. The present
 CC sequence represents a human mamaglobin N-terminal amino acid sequence
 CC fragment, which is used in an example from the present invention.
 XX
 SQ Sequence 55 AA;
 Query Match 7.4%; Score 32; DB 22; Length 55;
 Best Local Similarity 100.0%; Pred. No. 3.1e-24;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 15 GSGMKETAARAFERQHMDSPDLGTDDDDKAMA 46
 DB 1 GSGMKETAARAFERQHMDSPDLGTDDDDKAMA 32
 RESULT 38
 AAY39952
 ID AAY39952 standard; Protein: 382 AA.
 XX
 AC AAY39952;
 XX
 DT 13-DEC-1999 (first entry)
 XX
 DE Gaussia luciferase fusion protein sequence.
 XX
 KW Luciferase; green fluorescent protein; GFP; screening assay; diagnosis;
 KW bioluminescence-generating system; toy; cosmetic; fairy dust; beverage;
 KW body paint; squirt gun; balloon; slimy play material; soap; toothpaste;
 KW fusion protein.
 XX
 OS Gaussia sp.
 XX
 PN WO9949019-A2.
 XX
 PD 30-SEP-1999.
 XX
 PF 26-MAR-1999; 99WO-US06698.
 XX
 PR 27-MAR-1998; 98US-0079624.
 PR 15-JUN-1998; 98US-0089367.
 PR 01-OCT-1998; 98US-0102939.
 XX
 PA (PROL-) PROLUME LTD.
 PA (BRYA/) BRYAN B J.
 XX
 PI Bryan BJ, Szent-Gyorgyi C;
 XX
 DR WPI; 1999-580443/49.
 DR N-PSDB; AAZ27550.
 XX
 PT New isolated Renilla mulleri, Gaussia and Pleuromamma luciferase and
 PT Renilla and Ptilosarcus green fluorescent protein nucleic acids -
 XX
 PS Disclosure; Page 222-223; 233pp; English.
 XX
 CC This sequence represents a luciferase of the invention. The invention
 CC relates to Renilla mulleri, Gaussia and Pleuromamma luciferase and

CC Renilla and *Ptilosarcus* green fluorescent protein (GFP) nucleic acids and
 CC proteins. The luciferases and GFPs can be used in
 CC bioluminescence-generating systems, assays, screening methods, diagnostic
 CC method and articles of manufacture. They can be expressed using
 CC e.g. bacterial, yeast, fungal, plant, insect or animal cells. The Renilla
 CC mulleri, *Gaussia* and *Pleuromamma* luciferase or Renilla or *Ptilosarcus*
 CC GFPs can be used in e.g. toys, cosmetics, fountains, personal care items,
 CC fairy dust, beverages, soft drinks, foods, textile products, bubbles,
 CC balloons, personal items, dentifrices, soaps, body paints, bubble bath,
 CC ink or paper products. In particular, they can be used in e.g. squirt
 CC guns, pellet guns, finger paints, foot bags, greeting cards, slimy play
 CC material, clothing, bubble making toys, bath powders, cosmetics, body
 CC lotions, gels, body powders, body creams, bath powders, mouthwashes,
 CC soaps, body paints, bubble bath, inks, wrapping paper, gelatins, icings,
 CC frostings, greeting cards, beer, wine, champagne, soft drinks, ice cubes,
 CC ice, dry ice or fountains. The nucleic acids can also be used to produce
 CC transgenic fish and plants.

XX Sequence 382 AA;

Query Match 6.3%; Score 27; DB 20; Length 382;

Best Local Similarity 100.0%; Pred. No. 2.1e-18;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 GKKETAATAAKFERQHMDSPDLGTDDDDK 43

|||||

Db 171 GKKETAATAAKFERQHMDSPDLGTDDDDK 197

RESULT 39

AAE13383

ID AAE13383 standard; Protein; 382 AA.

XX AC AAE13383;

XX DT 12-FEB-2002 (first entry)

XX DE *Gaussia* species CBD-luciferase fusion protein.

XX KW Green fluorescent protein; GFP; bioluminescence generating system; toy;

XX LW luciferase; finger paint; slimy play material; fishing lure; sparkler;

XX KW doll; balloon; personal care item; cosmetic; bath powder; body cream;

XX KW tooth paste; mouth wash; soap; body paint; ornamental transgenic plant;

XX KW bubble bath; fountain; fairy dust; food; gelatin; icing; frosting; beer;

XX KW wine; champagne; milk; soft drink; ice cube; ice; foot bag; beverage;

XX KW dry ice; textile; clothing; paper product; greeting card; wrapping paper;

XX KW fusion protein.

XX OS *Gaussia* sp.

XX PN WO20016824-A2.

XX XX 20-SEP-2001.

XX PF 15-MAR-2001; 2001WO-US08277.

XX PR 15-MAR-2000; 2000US-189691P.

XX PA (PROL-) PROLUME LTD.

XX PA (BRYA/) BRYAN B J.

XX XX Bryan BJ, Szent-Gyorgyi C, Szczepaniak W;

XX WPI; 2002-010561/01.

XX DR N-PSDB; AAD22201.

XX XX

XX PT Nucleic acids encoding Renilla reniformis green fluorescent proteins,

XX useful in diagnostic bioluminescence procedures -

XX PS Disclosure; Page 163-164; 175pp; English.

XX XX

XX CC The patent discloses sea pansy (*Renilla reniformis*) green fluorescent

XX proteins (GFP) and their corresponding polynucleotides. The invention

CC also relates to sequences of the bioluminescence generating system
 CC (e.g. luciferase). *R. reniformis* GFP are used in diagnostic methods
 CC and in the production of novelty items such as toys (e.g. squirt gun,
 CC pellet gun, toy "Halloween" eggs, toy cigarettes, board/card game
 CC toys), finger paints, slimy play material, bubbles in bubble making
 CC toys, fishing lures, dolls, sparklers, magic wand toys, balloons,
 CC personal care item (e.g. cosmetic, bath powders, body creams, tooth
 CC pastes, mouth wash, soaps, body paints, bubble bath), ornamental
 CC transgenic plants, fountain, fairy dust, food (gelatins, icings,
 CC frosting, beers, wine, champagne, milk, soft drinks, ice cubes, ice,
 CC dry ice, beverage), textile (foot bag, clothing) and/or paper product
 CC (greeting cards, wrapping paper). The present sequence is CBD-luciferase
 CC fusion protein from *Gaussia* species.

XX Sequence 382 AA;

Query Match 6.3%; Score 27; DB 23; Length 382;

Best Local Similarity 100.0%; Pred. No. 2.1e-18;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 GKKETAATAAKFERQHMDSPDLGTDDDDK 43

|||||

Db 171 GKKETAATAAKFERQHMDSPDLGTDDDDK 197

RESULT 40

AAE13356

ID AAE13356 standard; peptide; 54 AA.

XX AC AAE13356;

XX DT 04-MAR-1994 (first entry)

XX XX MSP1EGF2A EGF3-like domain variant.

XX DE Epidermal growth factor 1; merozoite surface protein 1; malaria;

XX KW vaccine.

XX OS Plasmodium yoelii.

XX FH Key Location/Qualifiers

XX FT Cleavage-site 1

XX FT /note= "introduced to facilitate cleavage

XX FT from recombinant protein"

XX PN WO9317107-A.

XX PD 02-SEP-1993.

XX PF 22-FEB-1993; 93WO-GB00367.

XX PR 22-FEB-1992; 92GB-0003821.

XX XX (MEDI-) MEDICAL RES COUNCIL.

XX PI Blackman MJ, Chappel JA, Holder AA;

XX DR WPI; 1993-288413/36.

XX XX

XX PT Allelic variants of epidermal growth factor 1- or 2-like domains - of.

XX PT merozoite surface protein 1, produced recombinantly for malaria

XX PT vaccines

XX XX

XX PS Claim 2; Fig 2a; 35pp; English.

XX XX

XX CC The sequence is that of an allelic variant of a merozoite surface

XX protein-1 epidermal growth factor (EGF) 2-like domain. It may be

XX used alone or as part of a fusion protein of EGF-1-like and

XX EGF-2-like domains in vaccines against malaria. When expressed

XX recombinantly it is produced in a form indistinguishable from

XX CC that in the native protein.

XX XX

XX Sequence 54 AA;

Query Match 5.3%; Score 23; DB 14; Length 54;
Best Local Similarity 100.0%; Pred. No. 3.6e-15;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 409 KKITCECTKPDSPFLFDGIFCSS 431
|
Db 30 KKITCECTKPDSPFLFDGIFCSS 52

Search completed: May 12, 2003, 10:27:22
Job time : 62 secs

101	9	2.1	24	3	US-08-780-496-8	Sequence 8, Appl	174	7	1.6	133	4	US-09-556-877-304	Sequence 304, App
102	9	2.1	24	4	US-09-232-858B-27	Sequence 27, Appl	175	7	1.6	133	4	US-09-620-412C-304	Sequence 304, App
103	9	2.1	24	4	US-08-928-069-9	Sequence 9, Appl	176	7	1.6	167	4	US-09-507-323B-13	Sequence 13, Appl
104	9	2.1	53	2	US-08-651-818A-19	Sequence 19, Appl	177	7	1.6	191	4	US-09-561-500-13	Sequence 13, Appl
105	9	2.1	53	4	US-09-184-826-19	Sequence 19, Appl	178	7	1.6	191	4	US-09-561-108-13	Sequence 13, Appl
106	9	2.1	54	2	US-08-651-818A-23	Sequence 23, Appl	179	7	1.6	191	4	US-09-561-526-13	Sequence 13, Appl
107	9	2.1	54	4	US-08-184-826-23	Sequence 23, Appl	180	7	1.6	202	4	US-09-556-877-92	Sequence 92, Appl
108	9	2.1	131	1	US-08-441-629-6	Sequence 6, Appl	181	7	1.6	202	4	US-09-620-412C-92	Sequence 92, Appl
109	9	2.1	131	3	US-08-776-207-6	Sequence 6, Appl	182	7	1.6	206	4	US-09-183-861-26	Sequence 26, Appl
110	9	2.1	131	4	US-09-507-773-6	Sequence 6, Appl	183	7	1.6	206	4	US-09-022-765-26	Sequence 26, Appl
111	9	2.1	131	5	PCT-US95-09172-6	Sequence 6, Appl	184	7	1.6	210	1	US-08-234-783-2	Sequence 2, Appl
112	9	2.1	139	4	US-09-134-001C-4413	Sequence 4413, Ap	185	7	1.6	210	1	US-08-456-907-2	Sequence 2, Appl
113	9	2.1	254	3	US-09-167-434-7	Sequence 7, Appl	186	7	1.6	210	5	PCT-US95-05523-2	Sequence 2, Appl
114	9	2.1	323	3	US-08-853-755-7	Sequence 7, Appl	187	7	1.6	224	1	US-08-707-792A-6	Sequence 6, Appl
115	9	2.1	324	1	US-08-591-989-5	Sequence 5, Appl	188	7	1.6	224	1	US-08-707-792A-6	Sequence 6, Appl
116	9	2.1	462	1	US-08-480-604A-26	Sequence 26, Appl	189	7	1.6	224	4	US-09-605-785-825	Sequence 825, App
117	9	2.1	462	2	US-08-405-496A-26	Sequence 26, Appl	190	7	1.6	234	4	US-09-009-217-3	Sequence 3, Appl
118	9	2.1	462	4	US-08-915-136-26	Sequence 26, Appl	191	7	1.6	234	4	US-09-009-656-3	Sequence 3, Appl
119	8	1.9	17	2	US-08-600-783-13	Sequence 13, Appl	192	7	1.6	235	4	US-09-009-217-6	Sequence 6, Appl
120	8	1.9	60	1	US-08-255-457-1	Sequence 1, Appl	193	7	1.6	235	4	US-09-009-656-6	Sequence 6, Appl
121	8	1.9	60	2	US-09-115-032-1	Sequence 1, Appl	194	7	1.6	236	4	US-09-009-217-7	Sequence 7, Appl
122	8	1.9	60	5	PCT-US95-05772-1	Sequence 1, Appl	195	7	1.6	236	4	US-09-009-656-7	Sequence 7, Appl
123	8	1.9	61	2	US-08-359-222-2	Sequence 2, Appl	196	7	1.6	254	2	US-08-207-481-20	Sequence 20, Appl
124	8	1.9	61	3	US-09-094-563-2	Sequence 2, Appl	197	7	1.6	254	5	PCT-US95-02689-20	Sequence 20, Appl
125	8	1.9	73	4	US-09-366-887A-24	Sequence 24, Appl	198	7	1.6	255	2	US-08-207-481-26	Sequence 26, Appl
126	8	1.9	98	4	US-09-366-887A-6	Sequence 6, Appl	199	7	1.6	255	5	PCT-US95-02689-26	Sequence 26, Appl
127	8	1.9	98	4	US-09-366-887A-7	Sequence 7, Appl	200	7	1.6	267	4	US-09-643-597-352	Sequence 352, App
128	8	1.9	391	1	US-08-589-981-2	Sequence 2, Appl	201	7	1.6	267	4	US-09-643-597-354	Sequence 354, App
129	8	1.9	391	4	US-09-457-568-4	Sequence 4, Appl	202	7	1.6	304	4	US-09-605-785-835	Sequence 835, App
130	8	1.9	391	4	US-09-457-646-4	Sequence 4, Appl	203	7	1.6	321	4	US-09-582-660-5	Sequence 5, Appl
131	7	1.6	7	1	US-08-100-744-7	Sequence 7, Appl	204	7	1.6	332	4	US-08-818-112-53	Sequence 53, Appl
132	7	1.6	7	1	US-08-284-784-7	Sequence 7, Appl	205	7	1.6	332	4	US-08-818-111-53	Sequence 53, Appl
133	7	1.6	7	1	US-08-707-793A-7	Sequence 7, Appl	206	7	1.6	332	4	US-09-056-556-53	Sequence 53, Appl
134	7	1.6	7	1	US-08-707-792A-7	Sequence 7, Appl	207	7	1.6	332	4	US-09-072-596-53	Sequence 53, Appl
135	7	1.6	7	2	US-08-854-811-7	Sequence 7, Appl	208	7	1.6	344	3	US-08-733-837B-6	Sequence 6, Appl
136	7	1.6	8	4	US-09-072-993C-6	Sequence 6, Appl	209	7	1.6	363	3	US-08-881-771A-4	Sequence 4, Appl
137	7	1.6	8	4	US-09-072-993C-8	Sequence 8, Appl	210	7	1.6	378	1	US-08-707-793A-4	Sequence 4, Appl
138	7	1.6	8	4	US-09-326-529-16	Sequence 16, Appl	211	7	1.6	378	1	US-08-707-792A-4	Sequence 4, Appl
139	7	1.6	8	5	PCT-US95-04439-3	Sequence 3, Appl	212	7	1.6	384	1	US-08-707-793A-5	Sequence 5, Appl
140	7	1.6	10	1	US-08-524-757-30	Sequence 30, Appl	213	7	1.6	384	1	US-08-707-792A-5	Sequence 5, Appl
141	7	1.6	10	1	US-08-524-757-31	Sequence 31, Appl	214	7	1.6	401	4	US-08-965-762-20	Sequence 20, Appl
142	7	1.6	10	4	US-08-556-978B-10	Sequence 10, Appl	215	7	1.6	401	4	US-09-911-927-20	Sequence 20, Appl
143	7	1.6	10	4	US-09-425-638A-129	Sequence 129, App	216	7	1.6	401	4	US-09-911-882-20	Sequence 20, Appl
144	7	1.6	10	4	US-09-543-004-129	Sequence 129, App	217	7	1.6	403	2	US-09-061-337-10	Sequence 10, Appl
145	7	1.6	13	1	US-08-290-919-13	Sequence 13, Appl	218	7	1.6	403	2	US-09-122-129-10	Sequence 10, Appl
146	7	1.6	13	1	US-08-290-919-15	Sequence 15, Appl	219	7	1.6	403	3	US-09-340-991-10	Sequence 10, Appl
147	7	1.6	13	1	US-08-482-577B-15	Sequence 15, Appl	220	7	1.6	403	4	US-08-974-609-10	Sequence 10, Appl
148	7	1.6	13	4	US-09-218-176-40	Sequence 40, Appl	221	7	1.6	403	4	US-09-549-098-10	Sequence 10, Appl
149	7	1.6	13	4	US-08-182-176-40	Sequence 40, Appl	222	7	1.6	419	4	US-09-134-001C-3441	Sequence 3441, Ap
150	7	1.6	14	1	US-08-556-978B-40	Sequence 40, Appl	223	7	1.6	424	4	US-09-333-593A-8	Sequence 8, Appl
151	7	1.6	14	1	US-08-226-264-7	Sequence 7, Appl	224	7	1.6	424	4	US-09-605-785-617	Sequence 617, App
152	7	1.6	14	1	US-08-226-264-19	Sequence 19, Appl	225	7	1.6	449	4	US-09-620-412C-349	Sequence 349, App
153	7	1.6	14	2	US-08-226-264-20	Sequence 20, Appl	226	7	1.6	487	4	US-09-620-412C-333	Sequence 333, App
154	7	1.6	16	1	US-08-448-418-100	Sequence 100, App	227	7	1.6	518	4	US-09-620-412C-333	Sequence 333, App
155	7	1.6	16	1	US-08-226-264-8	Sequence 8, Appl	228	7	1.6	525	4	US-09-556-877-196	Sequence 196, App
156	7	1.6	16	1	US-08-226-264-21	Sequence 21, Appl	229	7	1.6	525	4	US-09-620-412C-196	Sequence 196, App
157	7	1.6	16	2	US-08-207-481-6	Sequence 6, Appl	230	7	1.6	534	1	US-08-440-377A-5	Sequence 5, Appl
158	7	1.6	16	5	PCT-US95-02689-6	Sequence 6, Appl	231	7	1.6	534	2	US-08-687-852-5	Sequence 5, Appl
159	7	1.6	17	5	US-08-031-603A-19	Sequence 19, Appl	232	7	1.6	576	4	US-09-541-990A-1	Sequence 1, Appl
160	7	1.6	17	5	PCT-US94-01176-19	Sequence 19, Appl	233	7	1.6	583	4	US-09-620-412C-353	Sequence 353, App
161	7	1.6	24	4	US-09-507-323B-9	Sequence 9, Appl	234	7	1.6	585	4	US-09-620-412C-337	Sequence 337, App
162	7	1.6	37	4	US-09-153-586-4	Sequence 4, Appl	235	7	1.6	619	4	US-09-620-412C-309	Sequence 309, App
163	7	1.6	37	4	US-09-507-323B-10	Sequence 10, Appl	236	7	1.6	631	4	US-09-620-412C-325	Sequence 325, App
164	7	1.6	50	4	US-09-507-323B-11	Sequence 11, Appl	237	7	1.6	642	4	US-08-737-109-10	Sequence 10, Appl
165	7	1.6	89	4	US-09-507-323B-12	Sequence 12, Appl	238	7	1.6	646	4	US-09-198-723A-60	Sequence 60, Appl
166	7	1.6	94	4	US-09-556-877-89	Sequence 89, Appl	239	7	1.6	646	4	US-09-198-723A-63	Sequence 63, Appl
167	7	1.6	94	4	US-09-620-412C-89	Sequence 89, Appl	240	7	1.6	646	4	US-09-198-723A-66	Sequence 66, Appl
168	7	1.6	120	2	US-08-207-481-33	Sequence 33, Appl	241	7	1.6	646	4	US-09-198-723A-69	Sequence 69, Appl
169	7	1.6	120	5	PCT-US95-02689-35	Sequence 35, Appl	242	7	1.6	646	4	US-09-198-723A-72	Sequence 72, Appl
170	7	1.6	121	2	US-08-207-481-31	Sequence 31, Appl	243	7	1.6	646	4	US-09-620-412C-317	Sequence 317, App
171	7	1.6	121	5	PCT-US95-02689-33	Sequence 33, Appl	244	7	1.6	654	4	US-09-620-412C-341	Sequence 341, App
172	7	1.6	128	4	US-09-012-097A-22	Sequence 22, Appl	245	7	1.6	683	2	US-09-620-412C-357	Sequence 357, App
173	7	1.6	129	4	US-09-556-877-91	Sequence 91, Appl	246	7	1.6	685	3	US-08-878-989-1	Sequence 1, Appl
					Sequence 91, Appl							US-09-136-282-2	Sequence 2, Appl

247	7	1.6	685	4	US-09-272-796-1	Sequence 1, Appl	320	6	1.4	6	4	US-09-083-366-12	Sequence 12, Appl
248	7	1.6	685	4	US-09-505-744-2	Sequence 2, Appl	321	6	1.4	6	4	US-08-995-659-18	Sequence 18, Appl
249	7	1.6	691	4	US-09-620-412C-313	Sequence 313, App	322	6	1.4	6	4	US-09-142-334-26	Sequence 26, Appl
250	7	1.6	700	4	US-09-620-412C-345	Sequence 345, App	323	6	1.4	6	4	US-09-215-649A-18	Sequence 18, Appl
251	7	1.6	715	4	US-09-620-412C-321	Sequence 321, App	324	6	1.4	6	4	US-09-140-084-7	Sequence 7, Appl
252	7	1.6	715	4	US-09-620-412C-329	Sequence 329, App	325	6	1.4	6	4	US-09-382-950-4	Sequence 4, Appl
253	7	1.6	778	4	US-09-556-877-193	Sequence 193, App	326	6	1.4	6	4	US-09-382-950-4	Sequence 4, Appl
254	7	1.6	778	4	US-09-620-412C-193	Sequence 193, App	327	6	1.4	6	4	US-08-977-378-2	Sequence 2, Appl
255	7	1.6	780	1	US-08-485-621-2	Sequence 2, Appl	328	6	1.4	6	4	US-09-353-555-2	Sequence 2, Appl
256	7	1.6	780	1	US-08-973-831-2	Sequence 2, Appl	329	6	1.4	6	4	US-09-426-568A-6	Sequence 6, Appl
257	7	1.6	780	5	PCT-US96-09530A-2	Sequence 2, Appl	330	6	1.4	6	4	US-08-868-452-15	Sequence 15, Appl
258	7	1.6	821	4	US-09-556-877-195	Sequence 195, App	331	6	1.4	6	4	US-08-868-452-22	Sequence 22, Appl
259	7	1.6	821	4	US-09-620-412C-195	Sequence 195, App	332	6	1.4	6	4	US-09-000-094-47	Sequence 47, Appl
260	7	1.6	869	1	US-08-188-582-32	Sequence 32, Appl	333	6	1.4	6	4	US-09-058-483-7	Sequence 7, Appl
261	7	1.6	869	1	US-08-646-715-32	Sequence 32, Appl	334	6	1.4	6	4	US-09-140-201-8	Sequence 8, Appl
262	7	1.6	948	4	US-09-556-877-194	Sequence 194, App	335	6	1.4	6	4	US-09-039-780A-106	Sequence 106, App
263	7	1.6	948	4	US-09-620-412C-194	Sequence 194, App	336	6	1.4	6	4	US-09-293-170-15	Sequence 15, Appl
264	7	1.6	1313	2	US-08-244-537-2	Sequence 2, Appl	337	6	1.4	6	4	US-09-291-170A-14	Sequence 14, Appl
265	7	1.6	2311	4	US-08-934-386-9	Sequence 9, Appl	338	6	1.4	6	4	US-09-577-780-18	Sequence 18, Appl
266	7	1.6	2311	3	US-08-417-089-6	Sequence 6, Appl	339	6	1.4	6	4	US-09-724-297-7	Sequence 7, Appl
267	7	1.6	2325	4	US-08-695-651-6	Sequence 6, Appl	340	6	1.4	6	4	US-09-183-841-3	Sequence 3, Appl
268	7	1.6	2325	4	US-08-930-285-6	Sequence 6, Appl	341	6	1.4	6	4	US-09-367-309A-4	Sequence 4, Appl
269	7	1.6	2325	4	US-08-695-421-6	Sequence 6, Appl	342	6	1.4	6	4	US-09-724-884-14	Sequence 14, Appl
270	7	1.6	2325	4	US-08-697-828A-10	Sequence 10, Appl	343	6	1.4	6	4	US-09-529-279-6	Sequence 6, Appl
271	6	1.4	6	1	US-07-862-495-1	Sequence 1, Appl	344	6	1.4	6	5	PCT-US94-04361-55	Sequence 55, Appl
272	6	1.4	6	1	US-08-160-670A-6	Sequence 6, Appl	345	6	1.4	6	5	PCT-US95-02689-8	Sequence 8, Appl
273	6	1.4	6	1	US-08-140-222-1	Sequence 1, Appl	346	6	1.4	6	5	PCT-US95-04439-2	Sequence 2, Appl
274	6	1.4	6	1	US-08-131-365B-48	Sequence 48, Appl	347	6	1.4	7	1	US-08-160-670A-1	Sequence 1, Appl
275	6	1.4	6	1	US-07-807-529A-72	Sequence 72, Appl	348	6	1.4	7	1	US-08-176-500-20	Sequence 20, Appl
276	6	1.4	6	1	US-08-170-095B-15	Sequence 15, Appl	349	6	1.4	7	1	US-08-471-052A-20	Sequence 20, Appl
277	6	1.4	6	1	US-08-110-300A-25	Sequence 25, Appl	350	6	1.4	7	2	US-08-189-331-20	Sequence 20, Appl
278	6	1.4	6	1	US-08-396-866-15	Sequence 15, Appl	351	6	1.4	7	2	US-08-471-939-20	Sequence 20, Appl
279	6	1.4	6	1	US-08-290-919-14	Sequence 14, Appl	352	6	1.4	7	2	US-08-471-800-20	Sequence 20, Appl
280	6	1.4	6	1	US-08-469-486-37	Sequence 37, Appl	353	6	1.4	7	2	US-08-471-068-20	Sequence 20, Appl
281	6	1.4	6	1	US-08-469-486-47	Sequence 47, Appl	354	6	1.4	7	4	US-09-331-362-3	Sequence 3, Appl
282	6	1.4	6	1	US-08-460-343B-51	Sequence 51, Appl	355	6	1.4	7	4	US-08-865-468-9	Sequence 9, Appl
283	6	1.4	6	1	US-08-299-567-1	Sequence 1, Appl	356	6	1.4	8	1	US-08-102-757-3	Sequence 3, Appl
284	6	1.4	6	1	US-08-178-476A-12	Sequence 12, Appl	357	6	1.4	8	2	US-08-762-433-6	Sequence 6, Appl
285	6	1.4	6	1	US-08-398-028B-51	Sequence 51, Appl	358	6	1.4	8	2	US-09-001-219-6	Sequence 6, Appl
286	6	1.4	6	1	US-08-660-626-4	Sequence 4, Appl	359	6	1.4	8	2	US-08-481-658B-22	Sequence 22, Appl
287	6	1.4	6	1	US-08-434-705B-17	Sequence 17, Appl	360	6	1.4	8	2	US-08-477-504A-22	Sequence 22, Appl
288	6	1.4	6	1	US-08-470-837-15	Sequence 15, Appl	361	6	1.4	8	2	US-08-486-756A-22	Sequence 22, Appl
289	6	1.4	6	1	US-08-470-837-22	Sequence 22, Appl	362	6	1.4	8	2	US-08-485-862B-22	Sequence 22, Appl
290	6	1.4	6	2	US-08-694-579-8	Sequence 8, Appl	363	6	1.4	8	3	US-08-787-739-22	Sequence 22, Appl
291	6	1.4	6	2	US-08-207-481-8	Sequence 14, Appl	364	6	1.4	8	3	US-08-487-077A-22	Sequence 22, Appl
292	6	1.4	6	2	US-08-472-244-14	Sequence 14, Appl	365	6	1.4	8	3	US-08-816-977-11	Sequence 11, Appl
293	6	1.4	6	2	US-08-448-418-99	Sequence 99, Appl	366	6	1.4	8	3	US-08-485-863A-22	Sequence 22, Appl
294	6	1.4	6	2	US-08-504-265B-51	Sequence 51, Appl	367	6	1.4	8	4	US-08-938-595-4	Sequence 4, Appl
295	6	1.4	6	2	US-08-348-155-8	Sequence 8, Appl	368	6	1.4	8	4	US-08-485-049D-22	Sequence 22, Appl
296	6	1.4	6	2	US-08-623-833B-16	Sequence 16, Appl	369	6	1.4	8	4	US-08-727-153-4	Sequence 4, Appl
297	6	1.4	6	2	US-08-652-507-8	Sequence 8, Appl	370	6	1.4	8	4	US-08-875-533-57	Sequence 57, Appl
298	6	1.4	6	2	US-08-668-123-48	Sequence 48, Appl	371	6	1.4	8	4	US-08-556-978B-13	Sequence 13, Appl
299	6	1.4	6	2	US-08-469-658-37	Sequence 37, Appl	372	6	1.4	8	4	US-08-637-732A-40	Sequence 40, Appl
300	6	1.4	6	2	US-08-469-658-47	Sequence 47, Appl	373	6	1.4	8	4	US-08-817-787-31	Sequence 31, Appl
301	6	1.4	6	2	US-08-595-043A-7	Sequence 7, Appl	374	6	1.4	8	4	US-09-178-115-22	Sequence 22, Appl
302	6	1.4	6	2	US-08-846-021A-21	Sequence 21, Appl	375	6	1.4	8	4	US-09-177-776-22	Sequence 22, Appl
303	6	1.4	6	2	US-08-846-021A-29	Sequence 29, Appl	376	6	1.4	8	4	US-09-549-831-15	Sequence 15, Appl
304	6	1.4	6	2	US-08-886-642-25	Sequence 25, Appl	377	6	1.4	9	1	US-08-155-171B-4	Sequence 4, Appl
305	6	1.4	6	2	US-09-086-201-17	Sequence 17, Appl	378	6	1.4	9	1	US-08-100-744-12	Sequence 12, Appl
306	6	1.4	6	2	US-08-591-196-57	Sequence 57, Appl	379	6	1.4	9	1	US-08-284-784-12	Sequence 12, Appl
307	6	1.4	6	2	US-08-712-878-7	Sequence 7, Appl	380	6	1.4	9	1	US-08-459-064B-29	Sequence 29, Appl
308	6	1.4	6	3	US-08-996-139-18	Sequence 18, Appl	381	6	1.4	9	1	US-08-323-445A-2	Sequence 2, Appl
309	6	1.4	6	3	US-08-300-928C-87	Sequence 87, Appl	382	6	1.4	9	1	US-08-515-903A-2	Sequence 2, Appl
310	6	1.4	6	3	US-08-430-944D-87	Sequence 87, Appl	383	6	1.4	9	1	US-08-737-757-12	Sequence 12, Appl
311	6	1.4	6	3	US-08-481-435-42	Sequence 42, Appl	384	6	1.4	9	2	US-08-460-421A-29	Sequence 29, Appl
312	6	1.4	6	3	US-08-772-440-28	Sequence 28, Appl	385	6	1.4	9	2	US-08-854-811-12	Sequence 12, Appl
313	6	1.4	6	3	US-08-430-014-87	Sequence 87, Appl	386	6	1.4	9	2	US-08-435-998-4	Sequence 4, Appl
314	6	1.4	6	3	US-08-431-184-87	Sequence 87, Appl	387	6	1.4	9	2	US-08-732-398-6	Sequence 6, Appl
315	6	1.4	6	4	US-09-267-031-18	Sequence 18, Appl	388	6	1.4	9	3	US-09-143-470-19	Sequence 19, Appl
316	6	1.4	6	4	US-09-398-341-1	Sequence 1, Appl	389	6	1.4	9	3	US-09-254-567-1	Sequence 1, Appl
317	6	1.4	6	4	US-09-031-168-4	Sequence 4, Appl	390	6	1.4	9	3	US-08-895-707-9	Sequence 9, Appl
318	6	1.4	6	4	US-08-789-333F-99	Sequence 99, Appl	391	6	1.4	9	4	US-08-875-533-36	Sequence 36, Appl
319	6	1.4	6	4	US-08-761-483-8	Sequence 8, Appl	392	6	1.4	9	4	US-08-556-978B-102	Sequence 102, App

393	6	1.4	9	4	US-09-076-804-1	Sequence 1, Appli	466	6	1.4	14	6	5187077-17	Patent No. 5187077
394	6	1.4	9	4	US-09-149-727-17	Sequence 17, Appl	467	6	1.4	14	6	5427925-15	Patent No. 5427925
395	6	1.4	9	5	PCT-US95-12840-2	Sequence 2, Appli	468	6	1.4	15	1	US-08-021-603A-14	Sequence 14, Appl
396	6	1.4	10	1	US-07-807-529A-73	Sequence 73, Appl	469	6	1.4	15	1	US-08-469-886-48	Sequence 48, Appl
397	6	1.4	10	1	US-08-340-812-16	Sequence 16, Appl	470	6	1.4	15	2	US-08-467-603-53	Sequence 53, Appl
398	6	1.4	10	1	US-08-459-064B-16	Sequence 16, Appl	471	6	1.4	15	2	US-08-466-793-53	Sequence 53, Appl
399	6	1.4	10	1	US-08-737-757-10	Sequence 10, Appl	472	6	1.4	15	2	US-08-469-658-48	Sequence 48, Appl
400	6	1.4	10	1	US-08-628-665-14	Sequence 14, Appl	473	6	1.4	15	2	US-08-491-861A-53	Sequence 53, Appl
401	6	1.4	10	2	US-08-482-142-150	Sequence 150, App	474	6	1.4	15	4	US-09-148-712-4	Sequence 4, Appl
402	6	1.4	10	2	US-08-460-421A-16	Sequence 16, Appl	475	6	1.4	15	5	PCT-US94-01176-14	Sequence 14, Appl
403	6	1.4	10	2	US-08-717-169-13	Sequence 13, Appl	476	6	1.4	16	1	US-08-175-393A-2	Sequence 2, Appl
404	6	1.4	10	2	US-08-220-240A-3	Sequence 3, Appli	477	6	1.4	16	1	US-08-578-649-13	Sequence 13, Appl
405	6	1.4	10	2	US-08-478-572-150	Sequence 150, App	478	6	1.4	16	4	US-08-706-945D-145	Sequence 145, App
406	6	1.4	10	2	US-08-792-553-24	Sequence 24, Appl	479	6	1.4	16	4	US-09-039-780A-103	Sequence 103, App
407	6	1.4	10	3	US-08-300-928C-88	Sequence 88, Appl	480	6	1.4	16	4	US-09-039-780A-105	Sequence 105, App
408	6	1.4	10	3	US-08-430-944D-88	Sequence 88, Appl	481	6	1.4	16	4	US-09-367-309A-3	Sequence 3, Appli
409	6	1.4	10	3	US-08-430-014-88	Sequence 88, Appl	482	6	1.4	16	4	US-09-155-171B-37	Sequence 37, Appl
410	6	1.4	10	3	US-08-431-184-88	Sequence 88, Appl	483	6	1.4	17	1	US-08-155-171B-37	Sequence 37, Appl
411	6	1.4	10	4	US-08-163-919A-17	Sequence 17, Appl	484	6	1.4	17	1	US-08-323-445A-18	Sequence 18, Appl
412	6	1.4	10	4	US-08-639-234-9	Sequence 9, Appli	485	6	1.4	17	1	US-08-513-903A-18	Sequence 18, Appl
413	6	1.4	10	4	US-08-484-296-150	Sequence 150, App	486	6	1.4	17	2	US-08-435-998-37	Sequence 37, Appl
414	6	1.4	10	4	US-08-039-780A-101	Sequence 101, App	487	6	1.4	17	2	US-08-651-818A-22	Sequence 22, Appl
415	6	1.4	10	4	US-09-099-053-27	Sequence 27, Appl	488	6	1.4	17	3	US-09-136-421-10	Sequence 10, Appl
416	6	1.4	10	4	US-09-029-785-9	Sequence 9, Appli	489	6	1.4	17	4	US-09-184-826-22	Sequence 22, Appl
417	6	1.4	10	5	PCT-US93-00909-13	Sequence 13, Appl	490	6	1.4	17	5	PCT-US95-12840-18	Sequence 18, Appl
418	6	1.4	10	5	PCT-US94-14073-17	Sequence 17, Appl	491	6	1.4	18	1	US-08-145-006C-6	Sequence 6, Appli
419	6	1.4	10	5	PCT-US95-04439-9	Sequence 9, Appli	492	6	1.4	18	1	US-08-578-649-14	Sequence 14, Appl
420	6	1.4	11	1	US-08-133-804-11	Sequence 11, Appl	493	6	1.4	18	2	US-08-356-060A-47	Sequence 47, Appl
421	6	1.4	11	1	US-08-170-095B-6	Sequence 6, Appli	494	6	1.4	18	3	US-08-670-175-8	Sequence 8, Appli
422	6	1.4	11	1	US-08-396-866-6	Sequence 6, Appli	495	6	1.4	18	3	US-08-895-707-15	Sequence 15, Appl
423	6	1.4	11	1	US-08-461-838-11	Sequence 11, Appl	496	6	1.4	18	4	US-08-460-900C-47	Sequence 47, Appl
424	6	1.4	11	1	US-08-727-757-4	Sequence 4, Appli	497	6	1.4	18	4	US-08-460-900C-61	Sequence 61, Appl
425	6	1.4	11	2	US-08-461-866-11	Sequence 11, Appl	498	6	1.4	18	4	US-08-674-509B-47	Sequence 47, Appl
426	6	1.4	12	1	US-08-100-744-11	Sequence 11, Appl	499	6	1.4	18	4	US-08-954-698-47	Sequence 47, Appl
427	6	1.4	12	1	US-08-284-784-11	Sequence 11, Appl	500	6	1.4	18	4	US-08-957-874-47	Sequence 47, Appl
428	6	1.4	12	1	US-08-459-064B-35	Sequence 35, Appl	501	6	1.4	18	5	PCT-US94-00545-6	Sequence 6, Appli
429	6	1.4	12	2	US-08-460-421A-35	Sequence 35, Appl	502	6	1.4	19	1	US-08-192-300-13	Sequence 13, Appl
430	6	1.4	12	2	US-08-854-811-11	Sequence 11, Appl	503	6	1.4	19	3	US-09-167-434-15	Sequence 15, Appl
431	6	1.4	12	2	US-08-732-398-10	Sequence 10, Appl	504	6	1.4	19	3	US-08-853-755-15	Sequence 15, Appl
432	6	1.4	12	2	US-08-732-398-12	Sequence 12, Appl	505	6	1.4	19	4	US-08-129-722A-6	Sequence 6, Appli
433	6	1.4	12	4	US-09-331-362-10	Sequence 10, Appl	506	6	1.4	19	4	US-08-129-722A-8	Sequence 8, Appli
434	6	1.4	12	4	US-08-938-595-5	Sequence 5, Appli	507	6	1.4	20	1	US-08-416-336-4	Sequence 4, Appli
435	6	1.4	12	4	US-08-727-153-5	Sequence 5, Appli	508	6	1.4	20	2	US-08-543-020-12	Sequence 12, Appl
436	6	1.4	12	4	US-08-983-275-4	Sequence 4, Appli	509	6	1.4	20	3	US-08-542-051-21	Sequence 21, Appl
437	6	1.4	12	4	US-08-327-874A-4	Sequence 4, Appli	510	6	1.4	20	4	US-08-640-737-32	Sequence 32, Appl
438	6	1.4	12	4	US-09-266-498-7	Sequence 7, Appli	511	6	1.4	20	4	US-08-122-458D-7	Sequence 7, Appli
439	6	1.4	12	4	US-09-376-330-6	Sequence 6, Appli	512	6	1.4	21	1	US-07-908-455A-23	Sequence 23, Appl
440	6	1.4	12	5	PCT-US94-09700-4	Sequence 4, Appli	513	6	1.4	21	1	US-08-434-120-49	Sequence 49, Appl
441	6	1.4	13	1	US-08-170-095B-10	Sequence 10, Appl	514	6	1.4	21	1	US-08-014-153D-45	Sequence 45, Appl
442	6	1.4	13	1	US-08-396-866-10	Sequence 10, Appl	515	6	1.4	21	1	US-08-455-325-49	Sequence 49, Appl
443	6	1.4	13	1	US-08-571-843A-4	Sequence 4, Appli	516	6	1.4	21	4	US-09-071-710-41	Sequence 41, Appl
444	6	1.4	13	1	US-08-439-747A-22	Sequence 22, Appl	517	6	1.4	21	4	US-08-912-276-25	Sequence 25, Appl
445	6	1.4	13	2	US-08-359-222-1	Sequence 1, Appli	518	6	1.4	21	4	US-09-153-804-17	Sequence 17, Appl
446	6	1.4	13	2	US-08-440-409B-22	Sequence 22, Appl	519	6	1.4	21	4	US-08-944-483-76	Sequence 76, Appl
447	6	1.4	13	2	US-08-415-593-21	Sequence 21, Appl	520	6	1.4	21	4	US-09-525-397-41	Sequence 41, Appl
448	6	1.4	13	3	US-09-094-563-1	Sequence 1, Appli	521	6	1.4	21	4	US-09-266-462-1	Sequence 1, Appli
449	6	1.4	13	4	US-08-981-321-8	Sequence 8, Appli	522	6	1.4	21	4	US-09-266-462-3	Sequence 3, Appli
450	6	1.4	13	4	US-09-331-362-8	Sequence 8, Appli	523	6	1.4	21	4	US-09-115-737-49	Sequence 49, Appl
451	6	1.4	13	4	US-09-147-805-7	Sequence 7, Appli	524	6	1.4	21	4	US-09-566-876-16	Sequence 16, Appl
452	6	1.4	13	4	US-09-418-785-3	Sequence 3, Appli	525	6	1.4	21	4	US-09-049-698-51	Sequence 51, Appl
453	6	1.4	13	4	US-09-116-492A-32	Sequence 32, Appl	526	6	1.4	21	4	US-09-065-383-33	Sequence 33, Appl
454	6	1.4	13	6	5215889-5	Patent No. 5215889	527	6	1.4	21	4	US-09-237-269-21	Sequence 21, Appl
455	6	1.4	14	1	US-07-961-837-6	Sequence 6, Appli	528	6	1.4	21	4	US-09-276-600-11	Sequence 11, Appl
456	6	1.4	14	1	US-07-807-529A-76	Sequence 76, Appl	529	6	1.4	22	4	US-09-039-780A-102	Sequence 102, App
457	6	1.4	14	1	US-08-232-453A-15	Sequence 15, Appl	530	6	1.4	23	2	US-08-273-146-37	Sequence 37, Appl
458	6	1.4	14	2	US-08-467-603-58	Sequence 58, Appl	531	6	1.4	23	2	US-08-350-260A-42	Sequence 42, Appl
459	6	1.4	14	2	US-08-466-793-58	Sequence 58, Appl	532	6	1.4	24	1	US-08-021-603A-16	Sequence 16, Appl
460	6	1.4	14	2	US-08-491-861A-58	Sequence 58, Appl	533	6	1.4	24	4	US-09-556-877-221	Sequence 221, App
461	6	1.4	14	3	US-08-300-928C-91	Sequence 91, Appl	534	6	1.4	24	4	US-09-045-494-2	Sequence 2, Appli
462	6	1.4	14	3	US-08-430-944D-91	Sequence 91, Appl	535	6	1.4	24	4	US-09-620-412C-221	Sequence 221, App
463	6	1.4	14	3	US-08-430-014-91	Sequence 91, Appl	536	6	1.4	24	5	PCT-US94-01176-16	Sequence 16, Appl
464	6	1.4	14	3	US-08-956-307B-3	Sequence 3, Appli	537	6	1.4	25	4	US-09-039-780A-96	Sequence 96, Appl
465	6	1.4	14	3	US-08-431-184-91	Sequence 91, Appl	538	6	1.4	26	1	US-08-227-372-2	Sequence 2, Appli

539	6	1.4	26	2	US-08-657-641-12	Sequence 12, Appl	612	6	1.4	77	4	US-09-643-597-361	Sequence 361, App
540	6	1.4	26	3	US-08-470-397-2	Sequence 2, Appl	613	6	1.4	79	2	US-08-448-418-97	Sequence 97, Appl
541	6	1.4	26	3	US-08-895-707-21	Sequence 21, Appl	614	6	1.4	80	4	US-09-174-943-4	Sequence 4, Appl
542	6	1.4	26	5	PCT-US94-07233-22	Sequence 12, Appl	615	6	1.4	81	2	US-08-717-169-15	Sequence 15, Appl
543	6	1.4	27	4	US-09-411-329C-22	Sequence 22, Appl	616	6	1.4	81	2	US-09-025-151-26	Sequence 26, Appl
544	6	1.4	27	4	US-09-411-329C-26	Sequence 26, Appl	617	6	1.4	85	3	US-08-772-440-32	Sequence 32, Appl
545	6	1.4	28	1	US-08-664-449-37	Sequence 37, Appl	618	6	1.4	85	3	US-09-256-976-82	Sequence 82, Appl
546	6	1.4	28	3	US-08-978-741-14	Sequence 14, Appl	619	6	1.4	87	1	US-08-160-670A-7	Sequence 7, Appl
547	6	1.4	28	4	US-09-333-729A-14	Sequence 14, Appl	620	6	1.4	88	1	US-08-308-086-11	Sequence 11, Appl
548	6	1.4	29	1	US-08-393-985-33	Sequence 33, Appl	621	6	1.4	89	1	US-08-167-035-22	Sequence 22, Appl
549	6	1.4	29	5	PCT-US94-05150-37	Sequence 37, Appl	622	6	1.4	89	1	US-08-340-812-13	Sequence 13, Appl
550	6	1.4	30	4	US-09-039-780A-98	Sequence 98, Appl	623	6	1.4	89	1	US-08-208-887A-22	Sequence 22, Appl
551	6	1.4	30	4	US-09-039-780A-100	Sequence 100, App	624	6	1.4	89	1	US-08-459-064B-13	Sequence 13, Appl
552	6	1.4	30	4	US-09-039-780A-104	Sequence 104, App	625	6	1.4	89	2	US-08-479-078-16	Sequence 16, Appl
553	6	1.4	31	2	US-08-662-227-37	Sequence 37, Appl	626	6	1.4	89	2	US-08-460-421A-23	Sequence 13, Appl
554	6	1.4	31	4	US-09-017-947-37	Sequence 37, Appl	627	6	1.4	89	2	US-08-539-005-72	Sequence 22, Appl
555	6	1.4	32	1	US-08-323-445A-17	Sequence 17, Appl	628	6	1.4	89	2	US-08-717-169-7	Sequence 7, Appl
556	6	1.4	32	1	US-08-515-903A-17	Sequence 17, Appl	629	6	1.4	89	4	US-09-280-598-24	Sequence 24, Appl
557	6	1.4	32	3	US-08-737-336-1	Sequence 1, Appl	630	6	1.4	94	4	US-09-256-976-95	Sequence 95, Appl
558	6	1.4	32	4	US-08-411-760-7	Sequence 7, Appl	631	6	1.4	96	1	US-07-807-529A-39	Sequence 39, Appl
559	6	1.4	32	4	US-08-411-760-9	Sequence 9, Appl	632	6	1.4	96	3	US-08-430-944D-103	Sequence 103, App
560	6	1.4	32	4	US-08-411-760-10	Sequence 10, Appl	633	6	1.4	96	3	US-08-431-184-103	Sequence 103, App
561	6	1.4	32	5	PCT-US95-12840-17	Sequence 17, Appl	634	6	1.4	96	4	US-09-605-785-827	Sequence 827, App
562	6	1.4	34	2	US-08-468-558-8	Sequence 8, Appl	635	6	1.4	97	3	US-08-816-977-23	Sequence 23, Appl
563	6	1.4	34	2	US-08-468-558-9	Sequence 9, Appl	636	6	1.4	97	3	US-08-816-977-27	Sequence 27, Appl
564	6	1.4	34	4	US-08-676-444-8	Sequence 8, Appl	637	6	1.4	97	4	US-09-174-943-2	Sequence 4, Appl
565	6	1.4	34	4	US-08-676-444-9	Sequence 9, Appl	638	6	1.4	106	1	US-07-623-611-4	Sequence 7, Appl
566	6	1.4	34	4	US-09-314-268-118	Sequence 118, App	639	6	1.4	106	1	US-07-623-611-7	Sequence 7, Appl
567	6	1.4	37	4	US-09-039-780A-97	Sequence 97, Appl	640	6	1.4	106	1	US-08-290-919-11	Sequence 11, Appl
568	6	1.4	39	2	US-08-679-865-39	Sequence 39, Appl	641	6	1.4	106	4	US-09-113-977C-47	Sequence 47, Appl
569	6	1.4	39	2	US-08-680-876-39	Sequence 39, Appl	642	6	1.4	106	5	PCT-US91-09108-4	Sequence 4, Appl
570	6	1.4	39	4	US-09-263-975-39	Sequence 39, Appl	643	6	1.4	106	5	PCT-US91-09108-7	Sequence 7, Appl
571	6	1.4	39	4	US-09-450-072-78	Sequence 78, Appl	644	6	1.4	107	2	US-08-722-349-6	Sequence 6, Appl
572	6	1.4	39	4	US-09-351-348-78	Sequence 78, Appl	645	6	1.4	107	3	US-09-204-328-6	Sequence 6, Appl
573	6	1.4	40	2	US-08-273-146-39	Sequence 39, Appl	646	6	1.4	108	2	US-08-468-819-74	Sequence 74, Appl
574	6	1.4	42	1	US-08-664-449-39	Sequence 39, Appl	647	6	1.4	108	2	US-08-468-819-76	Sequence 76, Appl
575	6	1.4	42	4	US-09-309-382-12	Sequence 12, Appl	648	6	1.4	120	1	US-08-347-492B-2	Sequence 2, Appl
576	6	1.4	43	2	US-08-273-146-41	Sequence 41, Appl	649	6	1.4	120	2	US-08-798-143-2	Sequence 2, Appl
577	6	1.4	43	2	US-08-652-816A-43	Sequence 43, Appl	650	6	1.4	120	4	US-08-341-560B-2	Sequence 2, Appl
578	6	1.4	44	2	US-08-652-507-6	Sequence 6, Appl	651	6	1.4	120	4	US-08-353-940-2	Sequence 2, Appl
579	6	1.4	44	4	US-08-960-190A-34	Sequence 34, Appl	652	6	1.4	120	5	PCT-US93-03895-2	Sequence 2, Appl
580	6	1.4	44	4	US-09-039-780A-99	Sequence 99, Appl	653	6	1.4	120	5	PCT-US95-15484-2	Sequence 2, Appl
581	6	1.4	45	4	US-08-960-190A-33	Sequence 33, Appl	654	6	1.4	121	4	US-08-341-560B-4	Sequence 4, Appl
582	6	1.4	48	1	US-08-457-245-21	Sequence 21, Appl	655	6	1.4	121	4	US-08-353-940-4	Sequence 4, Appl
583	6	1.4	50	4	US-09-360-237-53	Sequence 53, Appl	656	6	1.4	121	5	PCT-US93-03895-4	Sequence 4, Appl
584	6	1.4	52	3	US-08-978-741-15	Sequence 15, Appl	657	6	1.4	122	3	US-08-956-307B-19	Sequence 19, Appl
585	6	1.4	52	4	US-09-333-729A-15	Sequence 15, Appl	658	6	1.4	124	1	US-08-434-705B-16	Sequence 16, Appl
586	6	1.4	55	1	US-08-466-033-273	Sequence 273, App	659	6	1.4	124	2	US-09-086-201-16	Sequence 16, Appl
587	6	1.4	55	2	US-08-444-733-273	Sequence 273, App	660	6	1.4	124	4	US-09-199-242E-3	Sequence 3, Appl
588	6	1.4	55	2	US-08-464-134-273	Sequence 273, App	661	6	1.4	124	4	US-09-199-242E-5	Sequence 5, Appl
589	6	1.4	55	2	US-08-461-361-273	Sequence 273, App	662	6	1.4	124	4	US-09-042-071-39	Sequence 39, Appl
590	6	1.4	55	2	US-08-485-910-273	Sequence 273, App	663	6	1.4	126	1	US-08-268-348A-12	Sequence 12, Appl
591	6	1.4	57	2	US-08-652-816A-41	Sequence 41, Appl	664	6	1.4	126	6	5171845-12	Patent No. 5171845
592	6	1.4	57	2	US-08-958-201-18	Sequence 18, Appl	665	6	1.4	127	3	US-08-467-023-188	Sequence 188, App
593	6	1.4	60	4	US-09-076-804-4	Sequence 4, Appl	666	6	1.4	127	3	US-08-467-023-189	Sequence 189, App
594	6	1.4	63	4	US-09-134-001C-2901	Sequence 2901, Ap	667	6	1.4	127	3	US-08-467-023-187	Sequence 187, App
595	6	1.4	66	3	US-08-782-480-26	Sequence 26, Appl	668	6	1.4	128	3	US-08-467-023-187	Sequence 4, Appl
596	6	1.4	66	4	US-08-954-211-26	Sequence 26, Appl	669	6	1.4	130	4	US-08-240-839-4	Sequence 4, Appl
597	6	1.4	66	4	US-09-005-167A-26	Sequence 26, Appl	670	6	1.4	131	4	US-09-367-309A-2	Sequence 2, Appl
598	6	1.4	66	4	US-09-176-741B-26	Sequence 26, Appl	671	6	1.4	133	1	US-08-268-348A-8	Sequence 8, Appl
599	6	1.4	67	2	US-08-142-551B-9	Sequence 9, Appl	672	6	1.4	133	1	US-08-268-348A-10	Sequence 10, Appl
600	6	1.4	67	3	US-09-268-070-3	Sequence 3, Appl	673	6	1.4	136	5	PCT-US95-07171-2	Sequence 2, Appl
601	6	1.4	68	3	US-09-268-070-1	Sequence 1, Appl	674	6	1.4	139	4	US-08-856-253-8	Sequence 8, Appl
602	6	1.4	68	4	US-09-174-943-6	Sequence 6, Appl	675	6	1.4	147	4	US-08-122-458D-9	Sequence 9, Appl
603	6	1.4	69	2	US-08-687-865A-21	Sequence 21, Appl	676	6	1.4	149	4	US-09-134-001C-5151	Sequence 5151, Ap
604	6	1.4	69	3	US-09-268-070-2	Sequence 2, Appl	677	6	1.4	151	1	US-08-321-358-2	Sequence 2, Appl
605	6	1.4	71	4	US-08-043-711-21	Sequence 21, Appl	678	6	1.4	157	4	US-08-875-533-46	Sequence 46, Appl
606	6	1.4	71	1	US-08-309-152-4	Sequence 4, Appl	679	6	1.4	157	4	US-09-347-801-6	Sequence 6, Appl
607	6	1.4	71	4	US-09-146-054-9	Sequence 9, Appl	680	6	1.4	158	4	US-08-875-533-47	Sequence 47, Appl
608	6	1.4	71	4	US-09-664-977A-9	Sequence 9, Appl	681	6	1.4	159	3	US-08-991-890-4	Sequence 4, Appl
609	6	1.4	72	4	US-09-042-071-38	Sequence 38, Appl	682	6	1.4	159	4	US-08-856-253-2	Sequence 2, Appl
610	6	1.4	76	3	US-08-956-307B-17	Sequence 17, Appl	683	6	1.4	160	4	US-09-355-700-59	Sequence 59, Appl
611	6	1.4	76	3	US-08-956-307B-18	Sequence 18, Appl	684	6	1.4	171	1	US-08-680-726A-76	Sequence 76, Appl

685	1.4	171	4	US-09-092-409-76	Sequence 76, Appl	758	1.4	248	4	US-09-136-389-5	Sequence 5, Appl
686	1.4	172	3	US-08-860-165-12	Sequence 12, Appl	759	1.4	248	4	US-08-341-560B-8	Sequence 8, Appl
687	1.4	172	3	US-08-860-165-14	Sequence 14, Appl	760	1.4	248	4	US-09-610-838-5	Sequence 5, Appl
688	1.4	172	4	US-09-359-382-12	Sequence 12, Appl	761	1.4	248	4	US-08-353-940-8	Sequence 8, Appl
689	1.4	172	4	US-09-359-382-14	Sequence 14, Appl	762	1.4	248	5	PCT-US92-09487-5	Sequence 5, Appl
690	1.4	173	1	US-08-133-977-10	Sequence 10, Appl	763	1.4	248	5	PCT-US93-03895-8	Sequence 8, Appl
691	1.4	177	4	US-09-058-483-4	Sequence 4, Appl	764	1.4	249	1	US-08-155-171B-32	Sequence 32, Appl
692	1.4	178	4	US-09-042-071-42	Sequence 42, Appl	765	1.4	249	2	US-08-632-514C-11	Sequence 11, Appl
693	1.4	178	4	US-09-183-841-2	Sequence 2, Appl	766	1.4	249	2	US-08-435-998-32	Sequence 32, Appl
694	1.4	180	3	US-08-772-440-31	Sequence 31, Appl	767	1.4	249	3	US-09-188-177-11	Sequence 11, Appl
695	1.4	184	3	US-09-023-082A-30	Sequence 30, Appl	768	1.4	250	4	US-08-341-560B-6	Sequence 6, Appl
696	1.4	185	4	US-09-058-483-10	Sequence 10, Appl	769	1.4	250	4	US-08-353-940-6	Sequence 6, Appl
697	1.4	191	3	US-08-928-361B-26	Sequence 26, Appl	770	1.4	250	5	PCT-US93-03895-6	Sequence 6, Appl
698	1.4	191	5	PCT-US95-07171-3	Sequence 3, Appl	771	1.4	251	1	US-08-397-633A-74	Sequence 74, Appl
699	1.4	195	4	US-09-042-071-40	Sequence 40, Appl	772	1.4	254	2	US-08-767-026-7	Sequence 7, Appl
700	1.4	195	4	US-08-849-635-131	Sequence 3, Appl	773	1.4	254	2	US-08-792-824-4	Sequence 4, Appl
701	1.4	198	1	US-08-642-255-131	Sequence 131, Appl	774	1.4	254	2	US-08-792-824-7	Sequence 7, Appl
702	1.4	198	1	US-08-397-633A-41	Sequence 41, Appl	775	1.4	254	2	US-08-792-824-10	Sequence 10, Appl
703	1.4	198	1	US-08-397-633A-41	Sequence 52, Appl	776	1.4	254	2	US-08-792-824-13	Sequence 13, Appl
704	1.4	203	4	US-09-766-055A-4	Sequence 4, Appl	777	1.4	254	4	US-09-004-731-89	Sequence 89, Appl
705	1.4	203	4	US-09-149-476-383	Sequence 383, Appl	778	1.4	254	4	US-09-004-731-92	Sequence 92, Appl
706	1.4	210	1	US-08-286-872-2	Sequence 2, Appl	779	1.4	254	4	US-08-749-699-89	Sequence 89, Appl
707	1.4	210	4	US-08-612-973-28	Sequence 28, Appl	780	1.4	254	4	US-08-749-699-92	Sequence 92, Appl
708	1.4	210	4	US-08-927-597-28	Sequence 28, Appl	781	1.4	254	4	US-09-004-729-89	Sequence 89, Appl
709	1.4	211	4	US-08-856-253-4	Sequence 4, Appl	782	1.4	254	4	US-09-004-729-92	Sequence 92, Appl
710	1.4	215	1	US-08-107-684B-10	Sequence 10, Appl	783	1.4	255	2	US-08-933-750C-7	Sequence 7, Appl
711	1.4	215	1	US-08-107-684B-11	Sequence 11, Appl	784	1.4	255	3	US-09-167-434-10	Sequence 10, Appl
712	1.4	216	4	US-09-077-955-31	Sequence 31, Appl	785	1.4	255	3	US-08-853-755-10	Sequence 10, Appl
713	1.4	219	1	US-08-621-081A-16	Sequence 16, Appl	786	1.4	255	4	US-09-234-613-7	Sequence 7, Appl
714	1.4	220	4	US-09-485-885-1	Sequence 1, Appl	787	1.4	256	4	US-09-077-955-30	Sequence 30, Appl
715	1.4	220	4	US-09-485-885-8	Sequence 8, Appl	788	1.4	256	4	US-08-913-159-6	Sequence 6, Appl
716	1.4	220	4	US-09-134-001C-5073	Sequence 5073, Appl	789	1.4	257	4	US-08-149-091-2	Sequence 2, Appl
717	1.4	221	1	US-08-621-081A-19	Sequence 19, Appl	790	1.4	258	1	US-08-637-761-2	Sequence 2, Appl
718	1.4	221	4	US-09-134-001C-4984	Sequence 4984, Appl	791	1.4	258	5	PCT-US94-12672-2	Sequence 2, Appl
719	1.4	223	4	US-09-272-342B-8	Sequence 8, Appl	792	1.4	260	4	US-08-495-484-7	Sequence 7, Appl
720	1.4	224	4	US-09-129-030-24	Sequence 24, Appl	793	1.4	261	1	US-08-160-670A-45	Sequence 45, Appl
721	1.4	226	4	US-08-865-468-1	Sequence 1, Appl	794	1.4	261	4	US-09-468-738A-29	Sequence 29, Appl
722	1.4	226	4	US-08-898-649-1	Sequence 1, Appl	795	1.4	261	4	US-09-940-019-29	Sequence 29, Appl
723	1.4	226	4	US-09-134-001C-5520	Sequence 5520, Appl	796	1.4	262	1	US-08-397-633A-73	Sequence 73, Appl
724	1.4	227	4	US-09-485-885-16	Sequence 16, Appl	797	1.4	263	1	US-08-160-670A-32	Sequence 32, Appl
725	1.4	227	4	US-09-485-885-19	Sequence 19, Appl	798	1.4	266	3	US-08-860-165-10	Sequence 10, Appl
726	1.4	229	4	US-09-327-984A-8	Sequence 8, Appl	799	1.4	266	4	US-09-359-382-10	Sequence 10, Appl
727	1.4	230	4	US-09-485-737B-102	Sequence 102, Appl	800	1.4	266	4	US-09-367-309A-1	Sequence 1, Appl
728	1.4	231	4	US-08-129-722A-2	Sequence 2, Appl	801	1.4	267	4	US-08-818-112-142	Sequence 142, Appl
729	1.4	232	4	US-08-327-874A-6	Sequence 6, Appl	802	1.4	267	4	US-08-818-111-137	Sequence 137, Appl
730	1.4	232	5	PCT-US94-09700-6	Sequence 6, Appl	803	1.4	267	4	US-09-056-556-142	Sequence 142, Appl
731	1.4	233	2	US-08-480-753-2	Sequence 2, Appl	804	1.4	267	4	US-09-485-737B-2	Sequence 2, Appl
732	1.4	233	3	US-09-041-889-9	Sequence 9, Appl	805	1.4	267	4	US-09-072-596-137	Sequence 137, Appl
733	1.4	233	3	US-08-837-058-9	Sequence 9, Appl	806	1.4	269	2	US-08-424-641B-3	Sequence 3, Appl
734	1.4	234	4	US-09-009-217-2	Sequence 2, Appl	807	1.4	269	2	US-08-820-980-3	Sequence 3, Appl
735	1.4	234	4	US-09-009-656-2	Sequence 2, Appl	808	1.4	269	2	US-08-826-439-3	Sequence 3, Appl
736	1.4	235	4	US-09-485-737B-93	Sequence 93, Appl	809	1.4	269	4	US-09-070-408-132	Sequence 132, Appl
737	1.4	239	4	US-09-485-885-12	Sequence 12, Appl	810	1.4	269	4	US-08-913-159-4	Sequence 4, Appl
738	1.4	240	4	US-09-485-737B-91	Sequence 91, Appl	811	1.4	270	4	US-09-532-856-7	Sequence 7, Appl
739	1.4	242	3	US-09-167-434-8	Sequence 8, Appl	812	1.4	272	4	US-09-101-135-8	Sequence 8, Appl
740	1.4	242	3	US-08-853-755-8	Sequence 8, Appl	813	1.4	273	4	US-09-485-885-4	Sequence 4, Appl
741	1.4	243	3	US-09-167-434-9	Sequence 9, Appl	814	1.4	274	4	US-09-532-856-8	Sequence 8, Appl
742	1.4	243	3	US-08-853-755-9	Sequence 9, Appl	815	1.4	278	4	US-09-485-885-21	Sequence 21, Appl
743	1.4	244	2	US-08-480-753-4	Sequence 4, Appl	816	1.4	279	4	US-09-134-001C-4707	Sequence 4707, Appl
744	1.4	244	4	US-09-045-494-4	Sequence 4, Appl	817	1.4	280	4	US-09-214-0950-119	Sequence 119, Appl
745	1.4	246	3	US-09-167-434-11	Sequence 11, Appl	818	1.4	281	1	US-08-397-633A-75	Sequence 75, Appl
746	1.4	246	3	US-08-853-755-11	Sequence 11, Appl	819	1.4	281	4	US-09-423-439-44	Sequence 44, Appl
747	1.4	247	2	US-08-956-267A-2	Sequence 2, Appl	820	1.4	281	4	US-09-184-658-40	Sequence 40, Appl
748	1.4	247	4	US-09-230-196-16	Sequence 16, Appl	821	1.4	284	3	US-09-386-642-54	Sequence 54, Appl
749	1.4	248	1	US-07-901-707-5	Sequence 5, Appl	822	1.4	284	4	US-09-318-661-4	Sequence 4, Appl
750	1.4	248	1	US-07-988-430-5	Sequence 5, Appl	823	1.4	285	4	US-09-360-017-1	Sequence 1, Appl
751	1.4	248	1	US-08-425-336-5	Sequence 5, Appl	824	1.4	286	4	US-08-397-633A-76	Sequence 76, Appl
752	1.4	248	1	US-08-378-761A-75	Sequence 75, Appl	825	1.4	287	4	US-09-272-342B-9	Sequence 9, Appl
753	1.4	248	1	US-08-485-286-75	Sequence 75, Appl	826	1.4	288	4	US-09-423-439-38	Sequence 38, Appl
754	1.4	248	1	US-08-488-113B-5	Sequence 5, Appl	827	1.4	288	4	US-09-134-001C-3333	Sequence 3333, Appl
755	1.4	248	1	US-08-477-484B-5	Sequence 5, Appl	828	1.4	288	4	US-09-386-642-13	Sequence 13, Appl
756	1.4	248	2	US-08-646-360-5	Sequence 5, Appl	829	1.4	289	3	US-09-184-658-63	Sequence 63, Appl
757	1.4	248	4	US-08-839-765-5	Sequence 5, Appl	830	1.4	289	4	US-09-386-642-14	Sequence 14, Appl

831	6	1.4	290	2	US-08-903-801-1	Sequence 1, Appl	904	6	1.4	349	4	US-09-011-769A-64	Sequence 64, Appl
832	6	1.4	291	1	US-09-295-055-1	Sequence 1, Appl	905	6	1.4	351	2	US-08-701-191A-3	Sequence 3, Appl
833	6	1.4	291	1	US-08-102-757-9	Sequence 9, Appl	906	6	1.4	352	1	US-08-395-507-1	Sequence 1, Appl
834	6	1.4	292	4	US-09-485-885-10	Sequence 10, Appl	907	6	1.4	354	2	US-08-216-894-4	Sequence 4, Appl
835	6	1.4	295	1	US-08-149-091-1	Sequence 1, Appl	908	6	1.4	354	4	US-09-115-746-4	Sequence 4, Appl
836	6	1.4	295	2	US-08-637-761-1	Sequence 1, Appl	909	6	1.4	357	4	US-09-605-785-826	Sequence 826, Appl
837	6	1.4	295	5	PCT-US94-12672-1	Sequence 1, Appl	910	6	1.4	362	1	US-08-395-507-2	Sequence 2, Appl
838	6	1.4	298	1	US-08-149-091-6	Sequence 6, Appl	911	6	1.4	364	4	US-09-457-040B-34	Sequence 34, Appl
839	6	1.4	298	2	US-08-637-761-6	Sequence 6, Appl	912	6	1.4	365	4	US-09-457-040B-35	Sequence 35, Appl
840	6	1.4	298	4	US-09-318-661-2	Sequence 2, Appl	913	6	1.4	366	4	US-09-210-843-2	Sequence 2, Appl
841	6	1.4	298	5	PCT-US94-12672-6	Sequence 6, Appl	914	6	1.4	366	4	US-09-718-692-4	Sequence 4, Appl
842	6	1.4	300	4	US-09-277-078-2	Sequence 2, Appl	915	6	1.4	366	4	US-09-718-852-4	Sequence 4, Appl
843	6	1.4	300	4	US-09-561-756-36	Sequence 36, Appl	916	6	1.4	366	4	US-09-718-852-4	Sequence 4, Appl
844	6	1.4	300	4	US-09-227-721-36	Sequence 36, Appl	917	6	1.4	368	1	US-08-211-942-17	Sequence 17, Appl
845	6	1.4	301	2	US-08-661-052-14	Sequence 14, Appl	918	6	1.4	368	4	US-09-433-428D-58	Sequence 58, Appl
846	6	1.4	301	4	US-09-188-082-14	Sequence 14, Appl	919	6	1.4	368	4	US-09-000-094-20	Sequence 20, Appl
847	6	1.4	301	4	US-09-364-088-14	Sequence 14, Appl	920	6	1.4	371	4	US-09-483-885-6	Sequence 6, Appl
848	6	1.4	301	4	US-09-102-716-14	Sequence 14, Appl	921	6	1.4	375	4	US-09-572-191-4	Sequence 4, Appl
849	6	1.4	302	2	US-08-203-532F-4	Sequence 4, Appl	922	6	1.4	375	4	US-09-000-094-22	Sequence 22, Appl
850	6	1.4	302	2	US-08-893-853-3	Sequence 3, Appl	923	6	1.4	375	4	US-09-723-262-4	Sequence 4, Appl
851	6	1.4	302	3	US-08-950-860-16	Sequence 16, Appl	924	6	1.4	375	4	US-09-723-219-4	Sequence 4, Appl
852	6	1.4	302	4	US-09-113-921-3	Sequence 3, Appl	925	6	1.4	376	3	US-08-751-512-8	Sequence 8, Appl
853	6	1.4	302	4	US-09-078-465-4	Sequence 4, Appl	926	6	1.4	378	2	US-08-707-237A-104	Sequence 104, Appl
854	6	1.4	302	5	PCT-US95-01882A-4	Sequence 4, Appl	927	6	1.4	378	3	US-08-642-246-26	Sequence 26, Appl
855	6	1.4	303	2	US-08-203-532F-2	Sequence 2, Appl	928	6	1.4	378	4	US-09-451-206-26	Sequence 26, Appl
856	6	1.4	303	3	US-09-078-465-2	Sequence 7, Appl	929	6	1.4	378	5	PCT-US96-06229-26	Sequence 26, Appl
857	6	1.4	303	4	US-09-078-465-2	Sequence 2, Appl	930	6	1.4	382	4	US-09-724-519-4	Sequence 4, Appl
858	6	1.4	303	5	PCT-US95-01882A-2	Sequence 2, Appl	931	6	1.4	382	4	US-09-592-037-4	Sequence 4, Appl
859	6	1.4	306	4	US-09-386-642-53	Sequence 53, Appl	932	6	1.4	383	4	US-09-485-885-23	Sequence 23, Appl
860	6	1.4	308	1	US-08-493-568-2	Sequence 2, Appl	933	6	1.4	384	4	US-09-374-038-4	Sequence 4, Appl
861	6	1.4	308	2	US-08-793-958-2	Sequence 2, Appl	934	6	1.4	384	4	US-09-658-179-4	Sequence 4, Appl
862	6	1.4	308	4	US-09-804-778A-2	Sequence 2, Appl	935	6	1.4	386	3	US-08-895-707-7	Sequence 7, Appl
863	6	1.4	308	4	US-09-804-778A-4	Sequence 4, Appl	936	6	1.4	388	2	US-08-759-581B-9	Sequence 9, Appl
864	6	1.4	309	4	US-09-079-029-9	Sequence 9, Appl	937	6	1.4	388	4	US-09-304-711-9	Sequence 9, Appl
865	6	1.4	310	4	US-09-079-029-11	Sequence 11, Appl	938	6	1.4	388	4	US-09-173-281-9	Sequence 9, Appl
866	6	1.4	312	4	US-09-079-029-10	Sequence 10, Appl	939	6	1.4	390	4	US-09-485-885-14	Sequence 14, Appl
867	6	1.4	315	2	US-08-484-956-91	Sequence 91, Appl	940	6	1.4	391	2	US-08-244-851A-10	Sequence 10, Appl
868	6	1.4	315	2	US-08-757-653-91	Sequence 91, Appl	941	6	1.4	391	2	US-08-389-011-23	Sequence 23, Appl
869	6	1.4	315	4	US-08-520-946-91	Sequence 91, Appl	942	6	1.4	391	3	US-08-403-917A-23	Sequence 23, Appl
870	6	1.4	315	4	US-09-386-653A-9	Sequence 9, Appl	943	6	1.4	391	4	US-09-348-952A-23	Sequence 23, Appl
871	6	1.4	319	3	US-08-526-136-14	Sequence 14, Appl	944	6	1.4	391	5	PCT-US95-15656-2	Sequence 2, Appl
872	6	1.4	319	4	US-09-386-642-12	Sequence 12, Appl	945	6	1.4	397	3	US-08-978-741-6	Sequence 6, Appl
873	6	1.4	320	2	US-08-757-653-163	Sequence 163, Appl	946	6	1.4	397	4	US-09-333-729A-7	Sequence 7, Appl
874	6	1.4	320	2	US-08-823-516-61	Sequence 61, Appl	947	6	1.4	397	4	US-08-327-874A-11	Sequence 11, Appl
875	6	1.4	320	3	US-08-759-038-102	Sequence 102, Appl	948	6	1.4	397	5	PCT-US94-09700-11	Sequence 11, Appl
876	6	1.4	320	3	US-08-758-314-102	Sequence 102, Appl	949	6	1.4	399	3	US-08-860-519-8	Sequence 8, Appl
877	6	1.4	322	1	US-08-097-828-1	Sequence 1, Appl	950	6	1.4	400	4	US-09-086-010-2	Sequence 2, Appl
878	6	1.4	322	1	US-08-097-828-3	Sequence 3, Appl	951	6	1.4	409	4	US-09-572-191-6	Sequence 6, Appl
879	6	1.4	322	1	US-08-480-756-1	Sequence 1, Appl	952	6	1.4	409	4	US-09-723-262-6	Sequence 6, Appl
880	6	1.4	322	1	US-08-480-756-3	Sequence 3, Appl	953	6	1.4	409	4	US-09-723-219-6	Sequence 6, Appl
881	6	1.4	322	2	US-08-462-403-1	Sequence 1, Appl	954	6	1.4	412	3	US-08-560-005-9	Sequence 9, Appl
882	6	1.4	322	2	US-08-462-403-3	Sequence 3, Appl	955	6	1.4	412	4	US-09-418-540-9	Sequence 9, Appl
883	6	1.4	322	5	PCT-US92-00331-1	Sequence 1, Appl	956	6	1.4	412	4	US-09-471-396-1	Sequence 1, Appl
884	6	1.4	322	5	PCT-US92-00331-3	Sequence 3, Appl	957	6	1.4	413	3	US-08-669-408B-10	Sequence 10, Appl
885	6	1.4	322	5	PCT-US93-10419-1	Sequence 1, Appl	958	6	1.4	413	4	US-09-374-038-2	Sequence 2, Appl
886	6	1.4	322	5	PCT-US93-10419-3	Sequence 3, Appl	959	6	1.4	413	4	US-09-658-179-2	Sequence 2, Appl
887	6	1.4	323	3	US-08-816-977-21	Sequence 21, Appl	960	6	1.4	414	5	PCT-US92-06840-2	Sequence 2, Appl
888	6	1.4	326	3	US-08-816-977-25	Sequence 25, Appl	961	6	1.4	415	4	US-09-100-193-2	Sequence 2, Appl
889	6	1.4	326	4	US-09-230-196-20	Sequence 20, Appl	962	6	1.4	420	2	US-08-845-998-8	Sequence 8, Appl
890	6	1.4	326	4	US-09-632-947B-1	Sequence 1, Appl	963	6	1.4	420	4	US-09-206-537-8	Sequence 8, Appl
891	6	1.4	327	4	US-09-386-629-8	Sequence 8, Appl	964	6	1.4	420	4	US-09-430-854-8	Sequence 8, Appl
892	6	1.4	328	4	US-09-386-642-11	Sequence 11, Appl	965	6	1.4	422	4	US-09-217-228-7	Sequence 7, Appl
893	6	1.4	331	4	US-09-217-228-6	Sequence 6, Appl	966	6	1.4	423	2	US-08-910-798-2	Sequence 2, Appl
894	6	1.4	334	4	US-09-523-849-31	Sequence 31, Appl	967	6	1.4	425	1	US-08-313-553-13	Sequence 13, Appl
895	6	1.4	335	4	US-09-095-117-2	Sequence 2, Appl	968	6	1.4	425	3	US-08-767-993-13	Sequence 13, Appl
896	6	1.4	335	4	US-09-095-117-4	Sequence 4, Appl	969	6	1.4	426	3	US-08-932-823A-2	Sequence 2, Appl
897	6	1.4	336	4	US-09-134-001C-4379	Sequence 4379, Appl	970	6	1.4	426	3	US-08-737-248-4	Sequence 4, Appl
898	6	1.4	338	2	US-08-447-430A-43	Sequence 43, Appl	971	6	1.4	427	2	US-08-846-021A-8	Sequence 8, Appl
899	6	1.4	338	2	US-08-602-359A-40	Sequence 40, Appl	972	6	1.4	430	4	US-09-149-476-707	Sequence 707, Appl
900	6	1.4	345	4	US-08-856-253-7	Sequence 7, Appl	973	6	1.4	432	4	US-09-326-203A-25	Sequence 25, Appl
901	6	1.4	347	4	US-09-134-001C-3752	Sequence 3752, Appl	974	6	1.4	433	1	US-08-417-492-2	Sequence 2, Appl
902	6	1.4	349	4	US-09-011-769A-47	Sequence 47, Appl	975	6	1.4	434	4	US-09-548-372D-53	Sequence 53, Appl
903	6	1.4	349	4	US-09-011-769A-60	Sequence 60, Appl	976	6	1.4	434	4	US-09-548-367D-53	Sequence 53, Appl

777 6 1.4 435 4 US-09-134-001C-4622 Sequence 4622, Ap
778 6 1.4 439 4 US-09-433-428D-57 Sequence 57, Appl
779 6 1.4 439 4 US-08-506-296B-67 Sequence 67, Appl
780 6 1.4 440 3 US-09-259-214-2 Sequence 2, Appl
781 6 1.4 440 4 US-09-318-528-2 Sequence 2, Appl
782 6 1.4 440 4 US-09-291-331-2 Sequence 2, Appl
783 6 1.4 442 4 US-08-506-296B-70 Sequence 70, Appl
784 6 1.4 443 4 US-08-506-296B-76 Sequence 76, Appl
785 6 1.4 447 4 US-09-370-253-10 Sequence 10, Appl
786 6 1.4 447 4 US-08-506-296B-73 Sequence 73, Appl
787 6 1.4 448 4 US-09-461-474-8 Sequence 8, Appl
788 6 1.4 449 2 US-08-327-394-2 Sequence 2, Appl
789 6 1.4 459 4 US-09-548-372B-32 Sequence 32, Appl
790 6 1.4 459 4 US-09-548-367D-32 Sequence 32, Appl
791 6 1.4 464 2 US-08-553-619B-9 Sequence 9, Appl
792 6 1.4 465 4 US-09-000-094-24 Sequence 24, Appl
793 6 1.4 474 2 US-08-907-706-6 Sequence 6, Appl
794 6 1.4 474 2 US-09-096-982-8 Sequence 8, Appl
795 6 1.4 474 2 US-08-553-650A-8 Sequence 8, Appl
796 6 1.4 474 4 US-09-556-877-90 Sequence 90, Appl
797 6 1.4 474 4 US-09-620-412C-90 Sequence 90, Appl
798 6 1.4 477 3 US-08-772-270A-13 Sequence 13, Appl
799 6 1.4 479 1 US-08-397-633A-78 Sequence 78, Appl
1000 6 1.4 480 5 PCT-US93-04102-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-195-705-2
; Sequence 2, Application US/08195705
; Patent No. 6420523
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra
; APPLICANT: Hui, George
; APPLICANT: Barr, Philip
; APPLICANT: Gibson, Helen
; TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMIDIUM
; TITLE OF INVENTION: FALCIPARUM VACCINE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Davis Hoxie Faithfull Hapgood
; STREET: 45 Rockefeller Pl.
; CITY: New York
; STATE: N.Y.
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,705
; FILING DATE: 14-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jacobs, Seth H
; REGISTRATION NUMBER: 32140
; REFERENCE/DOCKET NUMBER: 11880A3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-757-2200
; TELEFAX: 212-586-1461
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 394 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; STRAIN: falciparum uganda palo alto (FUP)
; US-08-195-705-2
Query Match 53.6%; Score 231; DB 4; Length 394;
Best Local Similarity 100.0%; Pred. No. 4.5e-219;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 58 AISVTMDNIIISGSENEVDVYILKPLAGVYRSLLKQIEKNFTFNLDNLSRLKRRKY 117
Db 1 AISVTMDNIIISGSENEVDVYILKPLAGVYRSLLKQIEKNFTFNLDNLSRLKRRKY 60
QY 118 FLDVLESDLMQFKHISSENYIEDSFLLNSEQNTLLKSKYIKESVENDIKFAQEGIS 177
Db 61 FLDVLESDLMQFKHISSENYIEDSFLLNSEQNTLLKSKYIKESVENDIKFAQEGIS 120
QY 178 YYEKVLAKYKDDLESIRKVIKEEKEFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 237
Db 121 YYEKVLAKYKDDLESIRKVIKEEKEFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180
QY 238 YNNLVNKIDDYLYLNKAKINDCNVKEDEAHVKTFLSCLKAIDDKIDLFXN 288
Db 181 YNNLVNKIDDYLYLNKAKINDCNVKEDEAHVKTFLSCLKAIDDKIDLFXN 231
RESULT 2
US-08-195-705-4
; Sequence 4, Application US/08195705
; Patent No. 6420523
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra
; APPLICANT: Hui, George
; APPLICANT: Barr, Philip
; APPLICANT: Gibson, Helen
; TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMIDIUM
; TITLE OF INVENTION: FALCIPARUM VACCINE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Davis Hoxie Faithfull Hapgood
; STREET: 45 Rockefeller Pl.
; CITY: New York
; STATE: N.Y.
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,705
; FILING DATE: 14-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jacobs, Seth H
; REGISTRATION NUMBER: 32140
; REFERENCE/DOCKET NUMBER: 11880A3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-757-2200
; TELEFAX: 212-586-1461
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 394 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum

STRAIN: MAD
US-08-195-705-4

Query Match 31.3%; Score 135; DB 4; Length 394;
Best Local Similarity 100.0%; Pred. No. 1.2e-124; Indels 0; Gaps 0;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 LKSYKIKESVNDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTP 213
DB 97 LKSYKIKESVNDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTP 156

QY 214 SPKATDEOKKESKFLPFLNIETLYNNLVNKIDYLLINKAKINDCNVDEAHVITKL 273
DB 157 SPKATDEOKKESKFLPFLNIETLYNNLVNKIDYLLINKAKINDCNVDEAHVITKL 216

QY 274 SOLKAIDDKIDLFKN 288
DB 217 SOLKAIDDKIDLFKN 231

RESULT 3
US-08-290-919-4
; Sequence 4, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label=X
; OTHER INFORMATION: /note="X = M and N, or N"

US-08-290-919-4

Query Match 11.6%; Score 50; DB 1; Length 53;
Best Local Similarity 100.0%; Pred. No. 7.9e-42; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 382 PNPTCNENGGCDADATCTEEDSGSSRKKITCCTCKPDSYPLFDGIFCSS 431
DB 2 PNPTCNENGGCDADATCTEEDSGSSRKKITCCTCKPDSYPLFDGIFCSS 51

RESULT 4
US-09-741-243C-4
; Sequence 4, Application US/09741243C
; Patent No. 6399352
; GENERAL INFORMATION:
; APPLICANT: Crawford Jr., John Milton
; APPLICANT: Rice, John
; APPLICANT: Sevala, Veeresh
; APPLICANT: Stewart, Sandy
; TITLE OF INVENTION: NOVEL PLANT PORPHOBILINOGEN AND FUSION
; TITLE OF INVENTION: PROTEIN THEREOF
; FILE REFERENCE: 2022US
; CURRENT APPLICATION NUMBER: US/09/741,243C
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/171,785
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 167
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Thioredoxin functional fragment
US-09-741-243C-4

Query Match 11.6%; Score 50; DB 4; Length 167;
Best Local Similarity 100.0%; Pred. No. 2.2e-41; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDGDDDKAMADIGS 50
DB 116 MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDGDDDKAMADIGS 165

RESULT 5
US-09-513-442-2
; Sequence 2, Application US/09513442
; Patent No. 6387664
; GENERAL INFORMATION:
; APPLICANT: Ikemoto, Mitsushi
; TITLE OF INVENTION: SPARC Fusion Protein and a Process for Producing the
; TITLE OF INVENTION: Same
; FILE REFERENCE: HIRAKI-04218
; CURRENT APPLICATION NUMBER: US/09/513,442
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-513-442-2

Query Match 11.6%; Score 50; DB 4; Length 467;
Best Local Similarity 100.0%; Pred. No. 5.4e-41; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDGDDDKAMADIGS 50
DB 116 MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDGDDDKAMADIGS 165

RESULT 6

US-09-741-243C-2
; Sequence 2, Application US/09741243C
; Patent No. 6399352
; GENERAL INFORMATION:
; APPLICANT: Crawford Jr., John Milton
; APPLICANT: Rice, John
; APPLICANT: Sevala, Veeresh
; APPLICANT: Stewart, Sandy
; TITLE OF INVENTION: NOVEL PLANT PORPHOBILINOGEN AND FUSION
; TITLE OF INVENTION: PROTEIN THEREOF
; FILE REFERENCE: 2022US
; CURRENT APPLICATION NUMBER: US/09/741.243C
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/171,785
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A plant thioredoxin-porphobilinogen
; OTHER INFORMATION: synthase fusion protein
US-09-741-243C-2

Query Match 11.6%; Score 50; DB 4; Length 551;
Best Local Similarity 100.0%; Pred. No. 6.3e-41;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHSSGLVPRSGMKETAAAFERQHMDSPDLGTDGDDDDKAMADIGS 50
|||||
Db 116 MHHHHHSSGLVPRSGMKETAAAFERQHMDSPDLGTDGDDDDKAMADIGS 165
|||||

RESULT 7

US-09-626-589-3
; Sequence 3, Application US/09626589
; Patent No. 6326164
; GENERAL INFORMATION:
; APPLICANT: Rice, John
; APPLICANT: Kloti, Andreas
; APPLICANT: Crawford, John
; APPLICANT: Lanning, Beth
; APPLICANT: Stewart, Sandy
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: IDENTIFICATION OF MODULATORS OF DEOXYXYLULOSE 5-PHOSPHATE
; TITLE OF INVENTION: SYNTHASE ACTIVITY
; FILE REFERENCE: 2037 US
; CURRENT APPLICATION NUMBER: US/09/626.589
; CURRENT FILING DATE: 2000-07-27
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 824
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Residues 1-165 are from the thioredoxin sequence
; OTHER INFORMATION: found in the vector pET32 supplied by No. 6326164agen.
; OTHER INFORMATION: Residues 166-824 represent the tDXPS sequence from
; OTHER INFORMATION: Arabidopsis shown in SEQ ID NO:2.
US-09-626-589-3

Query Match 11.6%; Score 50; DB 4; Length 824;
Best Local Similarity 100.0%; Pred. No. 8.9e-41;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHSSGLVPRSGMKETAAAFERQHMDSPDLGTDGDDDDKAMADIGS 50
|||||
Db 116 MHHHHHSSGLVPRSGMKETAAAFERQHMDSPDLGTDGDDDDKAMADIGS 165
|||||

RESULT 8

US-08-290-919-2
; Sequence 2, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290.919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= X
; OTHER INFORMATION: /note= "X = M and N, or N"
US-08-290-919-2

Query Match 10.9%; Score 47; DB 1; Length 48;
Best Local Similarity 100.0%; Pred. No. 6.4e-39;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 339 ISQHOCVKKQCPENSGCFRHLDEREECKLLNLYKQEGDKCVENPNPT 385
|||||
Db 2 ISQHOCVKKQCPENSGCFRHLDEREECKLLNLYKQEGDKCVENPNPT 48
|||||

RESULT 9

US-08-683-007A-2
; Sequence 2, Application US/08683007A
; Patent No. 5858724
; GENERAL INFORMATION:
; APPLICANT: No. 5858724y, Robert E
; APPLICANT: Domanico, Michael

APPLICANT: Yeager, Keith
APPLICANT: Kroeker, Warren
TITLE OF INVENTION: Recombinant Rabbit Tissue Factor
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: USA
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/683,007A
FILING DATE: 16-JUL-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 740380.90040
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 408 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-683-007A-2

Query Match 10.7%; Score 46; DB 2; Length 408;
Best Local Similarity 100.0%; Pred. No. 4.1e-37;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHSSGLVPRGSMKETAARKFERQHMDSPDLGTDGDDKAMA 46
|||||
DB 116 MHHHHSSGLVPRGSMKETAARKFERQHMDSPDLGTDGDDKAMA 161

RESULT 10
US-08-195-705-5
Sequence 5, Application US/08195705
Patent No. 6420523
GENERAL INFORMATION:
APPLICANT: Chang, Sandra
APPLICANT: Hui, George
APPLICANT: Barr, Philip
APPLICANT: Gibson, Helen
TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMODIUM
TITLE OF INVENTION: FALCIPARUM VACCINE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Davis Hoxie Faithfull Hapgood
STREET: 45 Rockefeller Pl.
CITY: New York
STATE: N.Y.
COUNTRY: USA
ZIP: 10111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/195,705
FILING DATE: 14-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Jacobs, Seth H
REGISTRATION NUMBER: 32140
REFERENCE/DOCKET NUMBER: 11880A3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-757-2200
TELEFAX: 212-586-1461
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 377 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: C-terminal
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
STRAIN: WEL
US-08-195-705-5
Query Match 10.4%; Score 45; DB 4; Length 377;
Best Local Similarity 100.0%; Pred. No. 3.7e-36;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 353 SGCFRHLDERECKLLNLYKQEGDKCVENPNTCNENNGCDADA 397
|||||
DB 279 SGCFRHLDERECKLLNLYKQEGDKCVENPNTCNENNGCDADA 323
RESULT 11
US-08-290-919-1
Sequence 1, Application US/08290919
Patent No. 5720959
GENERAL INFORMATION:
APPLICANT: HOLDER, ANTHONY A.
APPLICANT: BLACKMAN, MICHAEL J.
APPLICANT: CHAPPEL, JONATHAN A.
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,919
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9203821.5
FILING DATE: 22-FEB-1992
PRIOR APPLICATION DATA: PCT/GB93/00367
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 212242/HCM/NJL/6BC8/
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids

```
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= X
; OTHER INFORMATION: /note= "X" = M and N, or "N"
US-08-290-919-1
Query Match 7.9%; Score 34; DB 1; Length 48;
Best Local Similarity 100.0%; Pred. No. 3.9e-26;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 352 NSGCFRHLDERECKCLLNKQEGDKCVENPNT 385
Db 15 NSGCFRHLDERECKCLLNKQEGDKCVENPNT 48
RESULT 12
US-08-290-919-12
; Sequence 12, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-290-919-12
Query Match 7.9%; Score 34; DB 1; Length 48;
Best Local Similarity 100.0%; Pred. No. 3.9e-26;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 352 NSGCFRHLDERECKCLLNKQEGDKCVENPNT 385
Db 15 NSGCFRHLDERECKCLLNKQEGDKCVENPNT 48
RESULT 13
US-09-166-966E-11
; Sequence 11, Application US/09166966E
; Patent No. 6407208
; GENERAL INFORMATION:
; APPLICANT: CHEN, DAVID CHANHAN
; APPLICANT: HU, NIEN-TAI
; APPLICANT: CHEN, YUN-JU
; APPLICANT: HSEU, TZONG-HSIUNG
; TITLE OF INVENTION: CHIMERIC PROTEINS WITH A CELLULOSE BINDING DOMAIN
; FILE REFERENCE: 32350-150960
; CURRENT APPLICATION NUMBER: US/09/166,966E
; CURRENT FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: TW 86114750
; PRIOR FILING DATE: 1997-10-08
; NUMBER OF SEQ ID NOS: 11
; SEQ ID NO 11
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; NAME/KEY: thioredoxin
; LOCATION: 1..109
; NAME/KEY: Cellulose-binding domain
; LOCATION: 162..197
; NAME/KEY: RGD
; LOCATION: 204..206
; OTHER INFORMATION: Combined amino acid sequence of thioredoxin,
; OTHER INFORMATION: artificial sequence, cellulose-binding domain, and
; OTHER INFORMATION: artificial sequence; the two artificial sequences
; OTHER INFORMATION: flanking the cellulose-binding domain contain restriction site
US-09-166-966E-11
Query Match 6.3%; Score 27; DB 4; Length 209;
Best Local Similarity 100.0%; Pred. No. 1.1e-18;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 20 ETAAAKFERQHMDSPDLGTDGDDDKAMA 46
Db 135 ETAAAKFERQHMDSPDLGTDGDDDKAMA 161
RESULT 14
US-09-277-716-22
; Sequence 22, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Artificial Sequence: fusion protein
; FEATURE:
; OTHER INFORMATION: Cellulose Binding Domain (CBD)-Gaussia1 luciferase fusion prote
```

US-09-277-716-22

Query Match 6.3%; Score 27; DB 4; Length 382;
Best Local Similarity 100.0%; Pred. No. 1.9e-18;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 GKKETAAAKFERQHMDSPDLGTDDDK 43
|||||
DB 171 GKKETAAAKFERQHMDSPDLGTDDDK 197

RESULT 15

US-09-609-161B-22
; Sequence 22, Application US/09609161B
; Patent No. 6436682

GENERAL INFORMATION:

; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.

; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HI
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS

; FILE REFERENCE: 24729-121B

; CURRENT APPLICATION NUMBER: US/09/609,161B

; CURRENT FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: 09/277,716

; PRIOR FILING DATE: 1999-03-26

; PRIOR APPLICATION NUMBER: 60/102,939

; PRIOR FILING DATE: 1998-10-01

; PRIOR APPLICATION NUMBER: 60/089,367

; PRIOR FILING DATE: 1998-06-15

; PRIOR APPLICATION NUMBER: 60/079,624

; PRIOR FILING DATE: 1998-03-27

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 22

; LENGTH: 382

; TYPE: PRT

; ORGANISM: Artificial Sequence: fusion protein

; FEATURE:

; OTHER INFORMATION: Cellulose Binding Domain (CBD)-Gaussia luciferase fusion protein

US-09-609-161B-22

Query Match 6.3%; Score 27; DB 4; Length 382;
Best Local Similarity 100.0%; Pred. No. 1.9e-18;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 GKKETAAAKFERQHMDSPDLGTDDDK 43
|||||
DB 171 GKKETAAAKFERQHMDSPDLGTDDDK 197

RESULT 16

US-09-166-966E-8

; Sequence 8, Application US/09166966E
; Patent No. 6407208

GENERAL INFORMATION:

; APPLICANT: CHEN, DAVID CHANHAN

; APPLICANT: HU, NIEN-TAI

; APPLICANT: CHEN, YUN-JU

; APPLICANT: HSEU, TZONG-HSIUNG

; TITLE OF INVENTION: CHIMERIC PROTEINS WITH A CELLULOSE BINDING DOMAIN

; FILE REFERENCE: 32350-150960

; CURRENT APPLICATION NUMBER: US/09/166,966E

; CURRENT FILING DATE: 1998-10-06

; PRIOR APPLICATION NUMBER: TW 86114750

; PRIOR FILING DATE: 1997-10-08

; NUMBER OF SEQ ID NOS: 11

; SEQ ID NO 8

; LENGTH: 159

; TYPE: PRT

; ORGANISM: Unknown

; FEATURE:

; NAME/KEY: thioredoxin

; LOCATION: 1..109

; OTHER INFORMATION: Combined amino acid of thioredoxin and an artificial sequence
US-09-166-966E-8

Query Match 5.6%; Score 24; DB 4; Length 159;
Best Local Similarity 100.0%; Pred. No. 7.7e-16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 ETAAAKFERQHMDSPDLGTDDDK 43
|||||
DB 135 ETAAAKFERQHMDSPDLGTDDDK 158

RESULT 17

US-08-195-705-3

; Sequence 3, Application US/08195705
; Patent No. 6420523

GENERAL INFORMATION:

; APPLICANT: Chang, Sandra

; APPLICANT: Hui, George

; APPLICANT: Barr, Philip

; APPLICANT: Gibson, Helen

; TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMODIUM

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Davis Hoxie Faithfull Hapgood

; STREET: 45 Rockefeller Pl.

; CITY: New York

; STATE: N.Y.

; COUNTRY: USA

; ZIP: 10111

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/195,705

; FILING DATE: 14-FEB-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Jacobs, Seth H

; REGISTRATION NUMBER: 32140

; REFERENCE/DOCKET NUMBER: 11880A3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-757-2200

; TELEFAX: 212-586-1461

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 375 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; HYPOTHETICAL: NO

; FRAGMENT TYPE: C-terminal

; ORIGINAL SOURCE:

; ORGANISM: Plasmodium falciparum

; STRAIN: K1

; US-08-195-705-3

Query Match 5.6%; Score 24; DB 4; Length 375;
Best Local Similarity 100.0%; Pred. No. 1.7e-15;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 NSGCFRHLDERECKCLLNKQEG 375
|||||
DB 276 NSGCFRHLDERECKCLLNKQEG 299

RESULT 18

US-08-290-919-3
; Sequence 3, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA: PCT/SB93/00367
; APPLICATION NUMBER: 212242/HCM/MJL/6BC8/
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= X
; OTHER INFORMATION: /note= "X" = M and N, or N"
US-08-290-919-3
Query Match 5.3%; Score 23; DB 1; Length 53;
Best Local Similarity 100.0%; Pred. No. 2.8e-15;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 409 KKITCECTKPDSPYPLFDGIFCSS 431
Db 29 KKITCECTKPDSPYPLFDGIFCSS 51
RESULT 19
US-08-470-837-24
; Sequence 24, Application US/08470837.
; Patent No. 5800811
; GENERAL INFORMATION:
; APPLICANT: Nimmi, Marcel E.
; APPLICANT: Hall, Frederick L.
; APPLICANT: Tuan, Tai-Lan
; APPLICANT: Wu, Lingtao
; APPLICANT: Cheung, David T.

US-08-470-837-24
; TITLE OF INVENTION: Transforming Growth Factor B Fusion
; TITLE OF INVENTION: and
; TITLE OF INVENTION: Their Use in Wound Healing
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 11150 Santa Monica Boulevard, Suite 400
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90025-3395
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,837
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharp, Janice A.
; REGISTRATION NUMBER: 34,051
; REFERENCE/DOCKET NUMBER: 30630-1US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310-445-1140
; TELEFAX: 310-445-9031
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-470-837-24
Query Match 3.5%; Score 15; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 KETAAKFERQHMD 33
Db 1 KETAAKFERQHMD 15
RESULT 20
US-08-639-806-1
; Sequence 1, Application US/08639806
; Patent No. 5817455
; GENERAL INFORMATION:
; APPLICANT: Raines, Ronald T
; TITLE OF INVENTION: Method for Enzyme Inactivation
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles and Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/639,806
; FILING DATE: 29-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/203,536
; FILING DATE: 01-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J

REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 70-399-9002-9
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-639-806-1

Query Match 3.5%; Score 15; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAKFERQHMD 33
|||||

Db 1 KETAAKFERQHMD 15

RESULT 21

US-08-868-452-24
Sequence 24, Application US/08868452C
Patent No. 6352972

GENERAL INFORMATION:
APPLICANT: Marcel E. Nimni
APPLICANT: Frederick L. Hall
APPLICANT: Lingtao Wu
APPLICANT: Bo Han
APPLICANT: Edwin Shors

TITLE OF INVENTION: BONE MORPHOGENETIC PROTEINS AND THEIR
TITLE OF INVENTION: USE IN BONE GROWTH
FILE REFERENCE: 17972-11
CURRENT APPLICATION NUMBER: US/08/868,452C
CURRENT FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 51
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 24
LENGTH: 15
TYPE: PRT
ORGANISM: Human
US-08-868-452-24

Query Match 3.5%; Score 15; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAKFERQHMD 33
|||||

Db 1 KETAAKFERQHMD 15

RESULT 22

US-08-373-134D-3
Sequence 3, Application US/08373134D
Patent No. 5780296

GENERAL INFORMATION:
APPLICANT: Kmiec, Eric
APPLICANT: Holloman, William

TITLE OF INVENTION: COMPOSITIONS AND METHODS TO PROMOTE
TITLE OF INVENTION: HOMOLOGOUS RECOMBINATION IN EUKARYOTIC CELLS AND ORGANISMS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA

ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/373,134D
FILING DATE: January 17, 1995

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Friebe, Thomas E.
REGISTRATION NUMBER: 29,258
REFERENCE/DOCKET NUMBER: 7991-007
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-373-134D-3

Query Match 3.5%; Score 15; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16
|||||

Db 5 HHHHHSSGLVPRGS 19

RESULT 23

US-09-114-637-3
Sequence 3, Application US/09114637
Patent No. 5945339

GENERAL INFORMATION:
APPLICANT: Kmiec, Eric
APPLICANT: Holloman, William

TITLE OF INVENTION: COMPOSITIONS AND METHODS TO PROMOTE
TITLE OF INVENTION: HOMOLOGOUS RECOMBINATION IN EUKARYOTIC CELLS AND ORGANISMS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA

ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/114,637
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/373,134
FILING DATE: January 17, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Friebe, Thomas E.
REGISTRATION NUMBER: 29,258
REFERENCE/DOCKET NUMBER: 7991-007
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-09-114-637-3

Query Match 3.5%; Score 15; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16
| | | | | | | | | | | | | | | | | | | |
Db 5 HHHHHSSGLVPRGS 19

RESULT 24

US-08-485-942A-99
; Sequence 99, Application US/08485942A
; Patent No. 6048837
; GENERAL INFORMATION:
; APPLICANT: JEFFREY M. FRIEDMAN, YIYING ZHANG, RICARDO PROENCA,
; APPLICANT: MARGHERITA MAFFEI, JEFFREY HALAAS, KETAN GAJIWALA, AND STEPHEN K. BURLE
; TITLE OF INVENTION: OB POLYPEPTIDE AS MODULATORS OF BODY WEIGHT (AS
; TITLE OF INVENTION: AMENDED)
; NUMBER OF SEQUENCES: 99
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485.942A
FILING DATE: JUNE 7, 1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/438,431
FILING DATE: May 10, 1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/347,563
FILING DATE: No. 6048837ember 30, 1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/292,345
FILING DATE: August 17, 1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-087 CIP 2F
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 99:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal His-tag
US-08-485-942A-99

Query Match 3.5%; Score 15; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.8e-08;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 HHHHHSSGLVPRGS 16
| | | | | | | | | | | | | | | | | | | |
Db 5 HHHHHSSGLVPRGS 19

RESULT 25

US-08-488-214A-99
; Sequence 99, Application US/08488214A
; Patent No. 6124439
; GENERAL INFORMATION:
; APPLICANT: JEFFREY M. FRIEDMAN, YIYING ZHANG, RICARDO PROENCA,
; APPLICANT: MARGHERITA MAFFEI, JEFFREY HALAAS, KETAN GAJIWALA, AND STEPHEN K. BU
; TITLE OF INVENTION: OB POLYPEPTIDE ANTIBODIES AND METHOD OF MAKING
; TITLE OF INVENTION: (AS AMENDED)
; NUMBER OF SEQUENCES: 99
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488.214A
FILING DATE: JUNE 7, 1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/438,431
FILING DATE: May 10, 1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/347,563
FILING DATE: No. 6124439ember 30, 1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/292,345
FILING DATE: August 17, 1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-087 CIP 2D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 99:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal His-tag
US-08-488-214A-99

Query Match 3.5%; Score 15; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.8e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16
| | | | | | | | | | | | | | | | | | | |
Db 5 HHHHHSSGLVPRGS 19

RESULT 26

US-08-438-431A-99

; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-432-871C-46

Query Match 3.5%; Score 15; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 9.2e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 HHHHHSSGLVPRGS 16
Db 5 HHHHHSSGLVPRGS 19

RESULT 29

US-08-712-878-6
; Sequence 6, Application US/08712878
; Patent No. 5985863
; GENERAL INFORMATION:
; APPLICANT: Su, Michael
; APPLICANT: Gu, Yong
; APPLICANT: Livingston, David J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DECREASING
; TITLE OF INVENTION: IGIF AND IFN-GAMMA PRODUCTION BY ADMINISTERING AN ICE
; TITLE OF INVENTION: INHIBITOR
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James F. Haley, Jr.
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/712,878
; FILING DATE: 12-SEP-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: VPI/96-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-712-878-6

Query Match 3.5%; Score 15; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 9.2e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16
Db 5 HHHHHSSGLVPRGS 19

RESULT 30

US-08-485-942A-98

; Sequence 98, Application US/08485942A
; Patent No. 6048837
; GENERAL INFORMATION:
; APPLICANT: JEFFREY M. FRIEDMAN, YIYING ZHANG, RICARDO PROENCA,
; APPLICANT: MARGHERITA MAFFEI, JEFFREY HALAAS, KETAN GAJIWALA, AND STEPHEN K. BU
; TITLE OF INVENTION: OB POLYPEPTIDE AS MODULATORS OF BODY WEIGHT (AS
; TITLE OF INVENTION: AMENDED)
; NUMBER OF SEQUENCES: 99
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,942A
; FILING DATE: JUNE 7, 1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/438,431
; FILING DATE: May 10, 1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/347,563
; FILING DATE: No. 6048837ember 30, 1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/292,345
; FILING DATE: August 17, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-087 CIP 2F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
US-08-485-942A-98

Query Match 3.5%; Score 15; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 9.2e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16
Db 5 HHHHHSSGLVPRGS 19

RESULT 31

US-08-488-214A-98
; Sequence 98, Application US/08488214A
; Patent No. 6124439

; GENERAL INFORMATION:
; APPLICANT: JEFFREY M. FRIEDMAN, YIYING ZHANG, RICARDO PROENCA,
; APPLICANT: MARGHERITA MAFFEI, JEFFREY HALAAS, KETAN GAJIWALA, AND STEPHEN K. BU
; TITLE OF INVENTION: OB POLYPEPTIDE ANTIBODIES AND METHOD OF MAKING
; TITLE OF INVENTION: (AS AMENDED)
; NUMBER OF SEQUENCES: 99
; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Klauber & Jackson
;; STREET: 411 Hackensack Avenue
;; CITY: Hackensack
;; STATE: New Jersey
;; COUNTRY: USA
;; ZIP: 07601
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/488,214A
;; FILING DATE: JUNE 7, 1995
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/438,431
;; FILING DATE: May 10, 1995
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/347,563
;; FILING DATE: NO. 6124439ember 30, 1994
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/438,431
;; FILING DATE: May 10, 1995
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/292,345
;; FILING DATE: August 17, 1994
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jackson Esq., David A.
;; REGISTRATION NUMBER: 26,742
;; REFERENCE/DOCKET NUMBER: 600-1-087 CIP 2D
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 201 487-5800
;; TELEFAX: 201 343-1684
;; TELEX: 133521
;; INFORMATION FOR SEQ ID NO: 98:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 21 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; FRAGMENT TYPE: N-terminal
;; US-08-488-214A-98

Query Match 3.5%; Score 15; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 9.2e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16
| | | | | | | | | | | | | | | | | | | | | |
Db 5 HHHHHSSGLVPRGS 19

RESULT 32
US-08-488-208A-98
; Sequence 98, Application US/08488208A
; Patent No. 6124448
; GENERAL INFORMATION:
; APPLICANT: THE ROCKEFELLER UNIVERSITY
; TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,211A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:

;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/488,208A
;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/485,943
;; FILING DATE: June 7, 1995
;; APPLICATION NUMBER: 08/438,431
;; FILING DATE: May 10, 1995
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/347,563
;; FILING DATE: NO. 6124448ember 30, 1994
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/292,345
;; FILING DATE: August 17, 1994
;; CLASSIFICATION: 514
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jackson Esq., David A.
;; REGISTRATION NUMBER: 26,742
;; REFERENCE/DOCKET NUMBER: 600-1-087 CIP2I
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 201 487-5800
;; TELEFAX: 201 343-1684
;; TELEX: 133521
;; INFORMATION FOR SEQ ID NO: 98:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 21 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; FRAGMENT TYPE: N-terminal
;; US-08-488-208A-98

Query Match 3.5%; Score 15; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 9.2e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16
| | | | | | | | | | | | | | | | | | | | | |
Db 5 HHHHHSSGLVPRGS 19

RESULT 33
US-08-483-211A-98
; Sequence 98, Application US/08483211A
; Patent No. 6309853
; GENERAL INFORMATION:
; APPLICANT: THE ROCKEFELLER UNIVERSITY
; TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,211A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/485,943
FILING DATE: June 7, 1995
APPLICATION NUMBER: 08/438,431
FILING DATE: May 10, 1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/347,563
FILING DATE: No. 6309853ember 30, 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/292,345
FILING DATE: August 17, 1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-087 CIP21
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 98:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal
US-08-483-211A-98

Query Match 3.5% Score 15; DB 4; Length 21;
Best Local Similarity 100.0%; Pred. No. 9.2e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16
| | | | | | | | | | | | | | | | | | | | | |
Db 5 HHHHHSSGLVPRGS 19

RESULT 34

US-08-488-223A-98
Sequence 98, Application US/08488223A
Patent No. 6350730

GENERAL INFORMATION:

APPLICANT: THE ROCKEFELLER UNIVERSITY
TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING NUCLEIC ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC USES THE

NUMBER OF SEQUENCES: 98

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488,223A

FILING DATE: 07-Jun-1995

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/485,943

FILING DATE: <Unknown>

APPLICATION NUMBER: 08/347,563

FILING DATE: No. 6350730ember 30, 1994

APPLICATION NUMBER: 08/292,345

FILING DATE: August 17, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-087 CIP21
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 98:

SEQUENCE CHARACTERISTICS:

LENGTH: 21 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: N-terminal

SEQUENCE DESCRIPTION: SEQ ID NO: 98:

US-08-488-223A-98

Query Match 3.5% Score 15; DB 4; Length 21;

Best Local Similarity 100.0%; Pred. No. 9.2e-08;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16

| | | | | | | | | | | | | | | | | | | | | |

Db 5 HHHHHSSGLVPRGS 19

RESULT 35

US-08-438-431A-98

Sequence 98, Application US/08438431A

Patent No. 6429290

GENERAL INFORMATION:

APPLICANT: JEFFREY M. FRIEDMAN, YIYING ZHANG, RICARDO PROENCA, MARGHERITA MAFEE

TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING NUCLEIC ACIDS AND

NUMBER OF SEQUENCES: 99

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/438,431A

FILING DATE: May 10, 1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/347,563

FILING DATE: No. 6429290ember 30, 1994

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/292,345

FILING DATE: August 17, 1994

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-087 CIP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 98:

SEQUENCE CHARACTERISTICS:

LENGTH: 21 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: N-terminal

US-08-438-431A-98

Query Match 3.5%; Score 15; DB 4; Length 21;
Best Local Similarity 100.0%; Pred. No. 9.2e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16
|||||
DB 5 HHHHHSSGLVPRGS 19

RESULT 36
US-09-270-956-46
; Sequence 46, Application US/09270956
; Patent No. 6451571
; GENERAL INFORMATION:
; APPLICANT: Loeb, Lawrence A.
; APPLICANT: Black, Margaret E.
; TITLE OF INVENTION: THYMIDINE KINASE MUTANTS
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/270,956
FILING DATE: 17-MAR-1999
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mcmasters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 240052.409C3
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-09-270-956-46
Query Match 3.5%; Score 15; DB 4; Length 21;
Best Local Similarity 100.0%; Pred. No. 9.2e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16
|||||
DB 5 HHHHHSSGLVPRGS 19

RESULT 37
US-08-761-483-7
; Sequence 7, Application US/08761483
; Patent No. 6204261
; GENERAL INFORMATION:
; APPLICANT: Batchelor, Mark J
; APPLICANT: Bebbington, David
; APPLICANT: Bemis, Guy W
; APPLICANT: Fridman, Wolf H
; APPLICANT: Gillespie, Roger J
; APPLICANT: Golec, Julian MC
; APPLICANT: Gu, Yong
; APPLICANT: Laufer, David J

APPLICANT: Livingston, David J
APPLICANT: Matharu, Saroop S
TITLE OF INVENTION: INHIBITORS OF INTERLEUKIN-1 BETA
TITLE OF INVENTION: CONVERTING ENZYME
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/761,483
FILING DATE: 06-DEC-1996
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr, James F
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: VPI/96-01CIP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS: single
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-761-483-7

Query Match 3.5%; Score 15; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 9.9e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16
|||||
DB 5 HHHHHSSGLVPRGS 19

RESULT 38
US-08-347-563A-12
; Sequence 12, Application US/08347563A
; Patent No. 5935810
; GENERAL INFORMATION:
; APPLICANT: THE ROCKEFELLER UNIVERSITY

TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING NUCLEIC
ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC USES TH.
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/347,563A
FILING DATE: No. 5935810ember 30, 1994
CLASSIFICATION: 514

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/292,345
;; FILING DATE: August 17, 1994
;; CLASSIFICATION: 514
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jackson Esq., David A.
;; REGISTRATION NUMBER: 26,742
;; REFERENCE/DOCKET NUMBER: 600-1-087 CIP
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 201 487-5800
;; TELEFAX: 201 343-1684
;; TELEX: 133521
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 43 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-347-563A-12

Query Match 3.5%; Score 15; DB 2; Length 43;
Best Local Similarity 100.0%; Pred. No. 1.7e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16
Db 5 HHHHHSSGLVPRGS 19

RESULT 39

US-08-485-942A-12
; Sequence 12, Application US/08485942A
; Patent No. 6048837
; GENERAL INFORMATION:
; APPLICANT: JEFFREY M. FRIEDMAN, YIYING ZHANG, RICARDO PROENCA,
; APPLICANT: MARGHERITA MAFFEI, JEFFREY HALAAS, KETAN GAJIWALA, AND STEPHEN K. BURLE
; TITLE OF INVENTION: OB POLYPEPTIDE AS MODULATORS OF BODY WEIGHT (AS
; TITLE OF INVENTION: AMENDED)
; NUMBER OF SEQUENCES: 99
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,942A
; FILING DATE: JUNE 7, 1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/438,431
; FILING DATE: May 10, 1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/347,563
; FILING DATE: August 17, 1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/292,345
; FILING DATE: August 17, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-087 CIP 2F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800

;; TELEFAX: 201 343-1684
;; TELEX: 133521
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 43 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-485-942A-12

Query Match 3.5%; Score 15; DB 3; Length 43;
Best Local Similarity 100.0%; Pred. No. 1.7e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16
Db 5 HHHHHSSGLVPRGS 19

RESULT 40

US-08-488-214A-12
; Sequence 12, Application US/08488214A
; Patent No. 6124439
; GENERAL INFORMATION:
; APPLICANT: JEFFREY M. FRIEDMAN, YIYING ZHANG, RICARDO PROENCA,
; APPLICANT: MARGHERITA MAFFEI, JEFFREY HALAAS, KETAN GAJIWALA, AND STEPHEN K. BU
; TITLE OF INVENTION: OB POLYPEPTIDE ANTIBODIES AND METHOD OF MAKING
; TITLE OF INVENTION: (AS AMENDED)
; NUMBER OF SEQUENCES: 99
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,214A
; FILING DATE: JUNE 7, 1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/438,431
; FILING DATE: May 10, 1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/347,563
; FILING DATE: No. 6124439ember 30, 1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/292,345
; FILING DATE: August 17, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-087 CIP 2D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-488-214A-12

Query Match 3.5%; Score 15; DB 3; Length 43;
Best Local Similarity 100.0%; Pred. No. 1.7e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16
| | | | | | | | | | | | | | | |
Db 5 HHHHHSSGLVPRGS 19

Search completed: May 12, 2003, 10:28:49
Job time : 33 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 10:28:19 ; Search time 22 Seconds
(without alignments)
1802.867 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 431

Sequence: 1 MHHHHHSSGLVPRSGMK.....TCECTKPDYPLFDGIFCSS 431

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 349150 seqs, 92025710 residues

Word size : 0

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PTCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/PTCTUS_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	374	86.8	394	9 US-09-978-756-3	Sequence 3, Appli
2	231	53.6	402	9 US-10-098-514-14	Sequence 14, Appl
3	227	52.7	383	9 US-10-098-514-4	Sequence 4, Appli
4	225	52.2	383	9 US-10-098-514-2	Sequence 2, Appli
5	60	13.9	95	10 US-09-134-333-2	Sequence 2, Appli
6	60	13.9	108	10 US-09-134-333-10	Sequence 10, Appl
7	60	13.9	116	10 US-09-134-333-5	Sequence 5, Appli
8	60	13.9	127	10 US-09-134-333-8	Sequence 8, Appli
9	50	11.6	516	10 US-09-804-626-4	Sequence 4, Appli
10	50	11.6	518	10 US-09-804-626-2	Sequence 2, Appli
11	50	11.6	824	9 US-10-046-583A-3	Sequence 3, Appli
12	46	10.7	96	9 US-09-978-756-1	Sequence 1, Appli
13	46	10.7	114	9 US-10-087-464-35	Sequence 35, Appl
14	46	10.7	201	9 US-09-681-938-1	Sequence 1, Appli
15	46	10.7	376	9 US-09-978-756-2	Sequence 2, Appli
16	46	10.7	376	9 US-10-087-464-34	Sequence 34, Appl
17	46	10.7	378	9 US-10-087-464-11	Sequence 11, Appl
18	46	10.7	1639	9 US-10-087-464-10	Sequence 10, Appl
19	45	10.4	355	9 US-10-012-896-1011	Sequence 1011, Ap

20	32	7.4	55	9	US-10-042-945-28	Sequence 28, Appl
21	32	7.4	55	10	US-09-757-417-28	Sequence 28, Appl
22	15	3.5	15	9	US-09-872-712-1	Sequence 1, Appli
23	15	3.5	18	10	US-09-905-831-1	Sequence 1, Appli
24	15	3.5	20	9	US-09-320-907B-5	Sequence 5, Appli
25	15	3.5	20	9	US-09-903-412-114	Sequence 114, App
26	15	3.5	20	10	US-09-096-749A-114	Sequence 5, Appli
27	15	3.5	21	9	US-10-060-275-5	Sequence 98, Appl
28	15	3.5	21	10	US-09-736-084-98	Sequence 12, Appl
29	15	3.5	43	10	US-09-736-084-12	Sequence 12, Appl
30	15	3.5	121	9	US-10-267-311-12	Sequence 8, Appli
31	15	3.5	124	9	US-09-981-286A-8	Sequence 19, Appl
32	15	3.5	149	9	US-09-876-348A-19	Sequence 23, Appl
33	15	3.5	149	9	US-09-876-348A-23	Sequence 27, Appl
34	15	3.5	149	9	US-09-876-348A-27	Sequence 31, Appl
35	15	3.5	149	9	US-09-876-348A-31	Sequence 35, Appl
36	15	3.5	149	9	US-09-876-348A-35	Sequence 39, Appl
37	15	3.5	149	9	US-09-876-348A-39	Sequence 19, Appl
38	15	3.5	149	9	US-09-876-796A-19	Sequence 23, Appl
39	15	3.5	149	9	US-09-876-796A-23	Sequence 27, Appl
40	15	3.5	149	9	US-09-876-796A-27	Sequence 31, Appl
41	15	3.5	149	9	US-09-876-796A-31	Sequence 35, Appl
42	15	3.5	149	9	US-09-876-796A-35	Sequence 39, Appl
43	15	3.5	149	9	US-09-876-796A-39	Sequence 2, Appli
44	15	3.5	153	10	US-09-798-338-2	Sequence 4, Appli
45	15	3.5	157	10	US-09-798-338-4	Sequence 6, Appli
46	15	3.5	163	10	US-09-798-338-6	Sequence 8, Appli
47	15	3.5	167	10	US-09-798-338-8	Sequence 4, Appli
48	15	3.5	170	9	US-10-068-059-4	Sequence 17, Appl
49	15	3.5	173	9	US-09-876-348A-17	Sequence 21, Appl
50	15	3.5	173	9	US-09-876-348A-21	Sequence 29, Appl
51	15	3.5	173	9	US-09-876-348A-29	Sequence 33, Appl
52	15	3.5	173	9	US-09-876-348A-33	Sequence 37, Appl
53	15	3.5	173	9	US-09-876-348A-37	Sequence 17, Appl
54	15	3.5	173	9	US-09-876-796A-17	Sequence 21, Appl
55	15	3.5	173	9	US-09-876-796A-21	Sequence 29, Appl
56	15	3.5	173	9	US-09-876-796A-33	Sequence 37, Appl
57	15	3.5	173	9	US-09-876-796A-37	Sequence 25, Appl
58	15	3.5	174	9	US-09-876-348A-25	Sequence 25, Appl
59	15	3.5	174	9	US-09-876-348A-25	Sequence 2, Appli
60	15	3.5	299	10	US-09-850-677A-2	Sequence 6, Appli
61	15	3.5	333	9	US-10-108-195-6	Sequence 16, Appl
62	15	3.5	333	9	US-09-849-243-16	Sequence 1, Appli
63	15	3.5	371	9	US-09-948-193-1	Sequence 2239, Ap
64	15	3.5	386	9	US-10-046-935-2239	Sequence 2239, Ap
65	15	3.5	391	9	US-10-146-502-2239	Sequence 121, App
66	15	3.5	391	9	US-08-979-847-121	Sequence 55, Appl
67	15	3.5	398	1	US-08-979-847-121	Sequence 8, Appli
68	15	3.5	409	9	US-10-267-311-55	Sequence 21, Appl
69	15	3.5	709	9	US-10-068-059-8	Sequence 17, Appl
70	15	3.5	746	9	US-10-068-059-6	Sequence 12, Appl
71	15	3.5	948	9	US-10-267-311-21	Sequence 13, Appl
72	14	3.2	948	9	US-09-849-243-12	Sequence 16, Appl
73	14	3.2	19	9	US-09-849-243-12	Sequence 9, Appli
74	9	2.1	13	9	US-10-108-195-13	Sequence 5, Appli
75	9	2.1	22	9	US-09-331-631A-16	Sequence 6, Appli
76	9	2.1	24	10	US-10-112-193-9	Sequence 9, Appli
77	9	2.1	24	10	US-09-934-465-8	Sequence 8, Appli
78	9	2.1	24	10	US-09-884-733-9	Sequence 9, Appli
79	9	2.1	24	12	US-10-080-455-5	Sequence 5, Appli
80	9	2.1	131	12	US-10-016-447-6	Sequence 40, Appl
81	9	2.1	135	10	US-09-804-615-40	Sequence 4, Appli
82	9	2.1	183	9	US-09-832-659-4	Sequence 2, Appli
83	9	2.1	183	9	US-09-832-658-2	Sequence 6, Appli
84	9	2.1	195	10	US-09-949-780-6	Sequence 11, Appl
85	9	2.1	600	9	US-09-930-125-11	Sequence 32, Appl
86	8	1.9	14	9	US-10-004-381-32	Sequence 12, Appl
87	8	1.9	31	9	US-09-996-069-12	Sequence 13, Appl
88	8	1.9	60	9	US-09-820-843A-13	Sequence 4, Appli
89	8	1.9	244	9	US-09-899-954B-4	Sequence 6, Appli
90	8	1.9	244	9	US-09-899-954B-6	Sequence 332, App
91	8	1.9	342	9	US-10-125-635A-332	Sequence 332, App
92	8	1.9	342	9	US-09-938-864-332	Sequence 332, App

93	8	1.9	342	9	US-09-785-019-332	Sequence 332, App	166	7	1.6	286	9	US-10-017-754-1878	Sequence 1878, App
94	8	1.9	391	8	US-08-902-572-2	Sequence 2, Appli	167	7	1.6	299	10	US-09-287-849-19	Sequence 19, App
95	8	1.9	410	9	US-10-125-635A-333	Sequence 333, App	168	7	1.6	299	10	US-09-735-705-354	Sequence 354, App
96	8	1.9	410	9	US-09-938-864-333	Sequence 333, App	169	7	1.6	299	10	US-09-850-716A-354	Sequence 354, App
97	8	1.9	410	9	US-09-785-019-333	Sequence 333, App	170	7	1.6	299	10	US-09-897-778-354	Sequence 354, App
98	8	1.9	544	10	US-09-864-761-48016	Sequence 48016, A	171	7	1.6	302	9	US-09-991-496-122	Sequence 122, App
99	8	1.9	591	9	US-10-125-635A-334	Sequence 334, App	172	7	1.6	302	10	US-09-874-923-122	Sequence 122, App
100	8	1.9	591	9	US-09-938-864-334	Sequence 334, App	173	7	1.6	304	9	US-10-012-896-835	Sequence 835, App
101	8	1.9	591	9	US-09-785-019-334	Sequence 334, App	174	7	1.6	304	9	US-09-895-793-835	Sequence 835, App
102	8	1.9	858	9	US-10-121-988-92	Sequence 92, Appli	175	7	1.6	304	9	US-09-895-814-835	Sequence 835, App
103	8	1.9	1356	10	US-09-801-368-306	Sequence 306, App	176	7	1.6	304	10	US-09-759-143-835	Sequence 835, App
104	7	1.6	7	9	US-09-872-712-30	Sequence 30, Appli	177	7	1.6	304	10	US-09-780-669-835	Sequence 835, App
105	7	1.6	8	9	US-10-156-634A-16	Sequence 16, Appli	178	7	1.6	304	10	US-09-822-827-835	Sequence 835, App
106	7	1.6	8	12	US-10-066-209-6	Sequence 6, Appli	179	7	1.6	311	9	US-10-121-988-93	Sequence 93, Appli
107	7	1.6	8	12	US-10-066-209-8	Sequence 8, Appli	180	7	1.6	311	9	US-09-736-457-1863	Sequence 1863, Ap
108	7	1.6	10	8	US-08-981-490B-5	Sequence 5, Appli	181	7	1.6	314	9	US-09-902-941-1863	Sequence 1863, Ap
109	7	1.6	10	10	US-09-065-040-26	Sequence 26, Appli	182	7	1.6	314	9	US-09-849-626-1863	Sequence 1863, Ap
110	7	1.6	13	10	US-09-823-266-5	Sequence 5, Appli	183	7	1.6	314	9	US-10-017-754-1863	Sequence 1863, Ap
111	7	1.6	37	9	US-09-858-580-4	Sequence 4, Appli	184	7	1.6	314	9	US-09-835-996A-25	Sequence 25, Appli
112	7	1.6	37	9	US-09-847-172-4	Sequence 4, Appli	185	7	1.6	330	12	US-10-007-693-122	Sequence 122, App
113	7	1.6	62	9	US-09-902-941-1880	Sequence 1880, Ap	186	7	1.6	331	10	US-09-287-849-4	Sequence 4, Appli
114	7	1.6	62	9	US-09-849-626-1880	Sequence 1880, Ap	187	7	1.6	336	9	US-09-987-107-44	Sequence 44, Appli
115	7	1.6	62	9	US-10-017-754-1880	Sequence 1880, Ap	188	7	1.6	337	9	US-09-987-107-46	Sequence 46, Appli
116	7	1.6	82	10	US-09-864-761-33313	Sequence 33313, A	189	7	1.6	344	9	US-10-025-380-1085	Sequence 1085, Ap
117	7	1.6	94	10	US-09-841-132-89	Sequence 89, Appli	190	7	1.6	344	10	US-09-922-217-1085	Sequence 1085, Ap
118	7	1.6	99	10	US-09-864-761-43778	Sequence 43778, A	191	7	1.6	344	10	US-09-833-263-1085	Sequence 1085, Ap
119	7	1.6	112	9	US-10-125-635A-345	Sequence 345, App	192	7	1.6	368	10	US-09-287-849-24	Sequence 24, Appli
120	7	1.6	112	9	US-09-938-864-345	Sequence 345, App	193	7	1.6	369	9	US-10-125-635A-346	Sequence 346, App
121	7	1.6	112	9	US-09-785-019-345	Sequence 345, App	194	7	1.6	369	9	US-09-938-864-346	Sequence 346, App
122	7	1.6	119	10	US-09-864-761-36209	Sequence 36209, A	195	7	1.6	369	9	US-09-785-019-346	Sequence 346, App
123	7	1.6	128	10	US-09-781-804-21	Sequence 21, Appli	196	7	1.6	377	9	US-10-121-988-96	Sequence 96, Appli
124	7	1.6	129	10	US-09-841-132-91	Sequence 91, Appli	197	7	1.6	380	10	US-09-134-333-12	Sequence 12, Appli
125	7	1.6	131	9	US-10-102-806-444	Sequence 444, Appli	198	7	1.6	380	10	US-09-134-333-13	Sequence 13, Appli
126	7	1.6	132	9	US-10-042-945-47	Sequence 47, Appli	199	7	1.6	384	9	US-09-902-941-1876	Sequence 1876, Ap
127	7	1.6	132	10	US-09-757-417-47	Sequence 47, Appli	200	7	1.6	384	9	US-09-849-626-1876	Sequence 1876, Ap
128	7	1.6	132	10	US-09-771-161A-158	Sequence 158, App	201	7	1.6	384	9	US-10-017-754-1876	Sequence 1876, Ap
129	7	1.6	133	10	US-09-841-132-304	Sequence 304, App	202	7	1.6	400	9	US-10-012-896-852	Sequence 852, App
130	7	1.6	164	10	US-09-798-789-2	Sequence 2, Appli	203	7	1.6	400	9	US-09-895-793-852	Sequence 852, App
131	7	1.6	164	10	US-09-981-289-2	Sequence 2, Appli	204	7	1.6	400	9	US-09-895-814-852	Sequence 852, App
132	7	1.6	188	9	US-10-125-635A-336	Sequence 336, App	205	7	1.6	400	9	US-10-026-021-5	Sequence 5, Appli
133	7	1.6	188	9	US-09-938-864-336	Sequence 336, App	206	7	1.6	400	10	US-09-759-143-852	Sequence 852, App
134	7	1.6	188	9	US-09-785-019-336	Sequence 336, App	207	7	1.6	400	10	US-09-780-669-852	Sequence 852, App
135	7	1.6	191	10	US-09-998-831-13	Sequence 13, Appli	208	7	1.6	400	10	US-09-822-827-852	Sequence 852, App
136	7	1.6	192	10	US-09-771-161A-159	Sequence 159, App	209	7	1.6	401	10	US-09-911-888-20	Sequence 20, Appli
137	7	1.6	193	9	US-09-991-496-119	Sequence 119, App	210	7	1.6	405	12	US-10-007-693-123	Sequence 123, App
138	7	1.6	193	10	US-09-874-923-119	Sequence 119, App	211	7	1.6	411	9	US-10-015-536-17	Sequence 17, Appli
139	7	1.6	202	10	US-09-841-132-92	Sequence 92, Appli	212	7	1.6	412	10	US-09-841-132-362	Sequence 362, App
140	7	1.6	206	9	US-09-991-496-26	Sequence 26, Appli	213	7	1.6	433	10	US-09-287-849-14	Sequence 14, Appli
141	7	1.6	206	10	US-09-874-923-26	Sequence 26, Appli	214	7	1.6	433	10	US-09-841-132-363	Sequence 363, App
142	7	1.6	207	10	US-09-771-161A-160	Sequence 160, App	215	7	1.6	449	9	US-10-012-896-617	Sequence 617, App
143	7	1.6	210	9	US-09-847-172-44	Sequence 44, Appli	216	7	1.6	449	9	US-09-895-793-617	Sequence 617, App
144	7	1.6	224	9	US-10-012-896-825	Sequence 825, App	217	7	1.6	449	9	US-09-895-814-617	Sequence 617, App
145	7	1.6	224	9	US-09-895-793-825	Sequence 825, App	218	7	1.6	449	10	US-09-759-143-617	Sequence 617, App
146	7	1.6	224	9	US-09-895-814-825	Sequence 825, App	219	7	1.6	449	10	US-09-780-669-617	Sequence 617, App
147	7	1.6	224	10	US-09-759-143-825	Sequence 825, App	220	7	1.6	449	10	US-09-822-827-617	Sequence 617, App
148	7	1.6	224	10	US-09-780-669-825	Sequence 825, App	221	7	1.6	487	10	US-09-841-132-349	Sequence 349, App
149	7	1.6	224	10	US-09-822-827-825	Sequence 825, App	222	7	1.6	511	10	US-09-949-780-2	Sequence 2, Appli
150	7	1.6	229	10	US-09-764-864-1437	Sequence 1437, Ap	223	7	1.6	518	10	US-09-841-132-333	Sequence 333, App
151	7	1.6	231	10	US-09-287-849-28	Sequence 28, Appli	224	7	1.6	521	10	US-09-925-297-474	Sequence 474, App
152	7	1.6	238	12	US-10-007-693-124	Sequence 124, App	225	7	1.6	525	10	US-09-841-132-196	Sequence 196, App
153	7	1.6	249	10	US-09-841-132-365	Sequence 365, App	226	7	1.6	529	9	US-09-924-400-324	Sequence 324, App
154	7	1.6	261	9	US-09-987-107-52	Sequence 52, Appli	227	7	1.6	529	10	US-09-810-936-324	Sequence 324, App
155	7	1.6	265	10	US-09-764-864-1006	Sequence 1006, Ap	228	7	1.6	583	10	US-09-841-132-353	Sequence 353, App
156	7	1.6	267	10	US-09-735-705-352	Sequence 352, App	229	7	1.6	585	9	US-09-895-793-978	Sequence 978, App
157	7	1.6	267	10	US-09-850-716A-352	Sequence 352, App	230	7	1.6	585	9	US-09-895-793-981	Sequence 981, App
158	7	1.6	267	10	US-09-897-778-352	Sequence 352, App	231	7	1.6	585	10	US-09-841-132-337	Sequence 337, App
159	7	1.6	273	9	US-09-736-457-1864	Sequence 1864, Ap	232	7	1.6	585	10	US-09-822-827-978	Sequence 978, App
160	7	1.6	273	9	US-09-902-941-1864	Sequence 1864, Ap	233	7	1.6	585	10	US-09-822-827-981	Sequence 981, App
161	7	1.6	273	9	US-09-849-626-1864	Sequence 1864, Ap	234	7	1.6	596	10	US-09-287-849-26	Sequence 26, Appli
162	7	1.6	273	9	US-10-017-754-1864	Sequence 1864, Ap	235	7	1.6	597	10	US-09-793-306-146	Sequence 146, App
163	7	1.6	281	10	US-09-835-996A-45	Sequence 45, Appli	236	7	1.6	598	12	US-10-007-693-140	Sequence 140, App
164	7	1.6	286	9	US-09-902-941-1878	Sequence 1878, Ap	237	7	1.6	600	10	US-09-287-849-22	Sequence 22, Appli
165	7	1.6	286	9	US-09-849-626-1878	Sequence 1878, Ap	238	7	1.6	619	10	US-09-841-132-309	Sequence 309, App

239	7	1.6	623	9	US-10-108-605-125	Sequence 125, App	312	6	1.4	6	10	US-09-280-030-61	Sequence 61, Appl
240	7	1.6	623	9	US-10-108-605-129	Sequence 129, App	313	6	1.4	6	10	US-09-771-956-11	Sequence 11, Appl
241	7	1.6	631	10	US-09-841-132-325	Sequence 325, App	314	6	1.4	6	10	US-09-822-485-30	Sequence 30, Appl
242	7	1.6	646	10	US-09-841-132-317	Sequence 317, App	315	6	1.4	6	10	US-09-760-008A-9	Sequence 9, Appli
243	7	1.6	654	10	US-09-841-132-341	Sequence 341, App	316	6	1.4	6	10	US-09-728-911-12	Sequence 12, Appl
244	7	1.6	660	12	US-10-007-693-139	Sequence 139, App	317	6	1.4	6	10	US-09-809-517A-11	Sequence 11, Appl
245	7	1.6	683	10	US-09-841-132-357	Sequence 357, App	318	6	1.4	6	10	US-09-211-691-14	Sequence 14, Appl
246	7	1.6	685	10	US-09-771-161A-249	Sequence 249, App	319	6	1.4	6	10	US-09-808-037-33	Sequence 33, Appl
247	7	1.6	685	10	US-09-771-161A-250	Sequence 250, App	320	6	1.4	6	10	US-09-997-900-6	Sequence 6, Appli
248	7	1.6	685	10	US-09-771-161A-251	Sequence 251, App	321	6	1.4	6	10	US-09-863-766A-9	Sequence 9, Appli
249	7	1.6	691	10	US-09-841-132-313	Sequence 313, App	322	6	1.4	6	10	US-09-888-615-149	Sequence 149, App
250	7	1.6	700	10	US-09-841-132-345	Sequence 345, App	323	6	1.4	6	10	US-09-875-519A-26	Sequence 26, Appl
251	7	1.6	704	9	US-10-121-988-94	Sequence 94, Appl	324	6	1.4	6	10	US-09-837-992-44	Sequence 44, Appl
252	7	1.6	710	10	US-09-287-849-16	Sequence 16, Appl	325	6	1.4	6	10	US-09-871-856-18	Sequence 18, Appl
253	7	1.6	715	10	US-09-841-132-321	Sequence 321, App	326	6	1.4	6	10	US-09-801-968-22	Sequence 22, Appl
254	7	1.6	715	10	US-09-841-132-329	Sequence 329, App	327	6	1.4	6	10	US-09-875-338-93	Sequence 93, Appl
255	7	1.6	729	10	US-09-287-849-2	Sequence 2, Appli	328	6	1.4	6	10	US-09-988-899-5	Sequence 5, Appli
256	7	1.6	778	10	US-09-841-132-193	Sequence 193, App	329	6	1.4	6	10	US-09-798-384-16	Sequence 16, Appl
257	7	1.6	801	9	US-09-895-793-982	Sequence 982, App	330	6	1.4	6	10	US-09-817-764-9	Sequence 9, Appli
258	7	1.6	801	9	US-09-822-827-982	Sequence 982, App	331	6	1.4	6	10	US-09-801-676A-2	Sequence 2, Appli
259	7	1.6	821	10	US-09-841-132-195	Sequence 195, App	332	6	1.4	6	10	US-09-916-940-99	Sequence 99, Appl
260	7	1.6	825	9	US-10-010-802-3	Sequence 3, Appli	333	6	1.4	6	10	US-09-780-933-14	Sequence 14, Appl
261	7	1.6	825	10	US-09-785-934-2	Sequence 2, Appli	334	6	1.4	6	10	US-09-900-530A-15	Sequence 15, Appl
262	7	1.6	825	10	US-09-921-667-16	Sequence 16, Appl	335	6	1.4	6	10	US-09-973-145-4	Sequence 4, Appli
263	7	1.6	825	10	US-09-904-245-2	Sequence 2, Appli	336	6	1.4	6	10	US-09-731-558-11	Sequence 11, Appl
264	7	1.6	842	10	US-09-771-161A-163	Sequence 163, App	337	6	1.4	6	10	US-09-822-698A-22	Sequence 22, Appl
265	7	1.6	856	10	US-09-287-849-12	Sequence 12, App	338	6	1.4	6	12	US-10-003-496-9	Sequence 9, Appli
266	7	1.6	904	9	US-10-043-487-382	Sequence 382, App	339	6	1.4	6	12	US-10-005-646-16	Sequence 16, Appl
267	7	1.6	926	9	US-09-991-496-128	Sequence 128, App	340	6	1.4	7	9	US-10-108-195-14	Sequence 14, Appl
268	7	1.6	948	10	US-09-841-132-194	Sequence 194, App	341	6	1.4	7	9	US-09-963-761B-5	Sequence 5, Appli
269	7	1.6	955	9	US-09-991-496-127	Sequence 127, App	342	6	1.4	7	9	US-10-047-991-9	Sequence 9, Appli
270	7	1.6	958	9	US-10-025-380-1087	Sequence 1087, Ap	343	6	1.4	7	10	US-09-809-517A-12	Sequence 12, Appl
271	7	1.6	958	10	US-09-922-217-1087	Sequence 1087, Ap	344	6	1.4	7	10	US-09-809-517A-13	Sequence 13, Appl
272	7	1.6	958	10	US-09-833-263-1087	Sequence 1087, Ap	345	6	1.4	8	9	US-10-116-273-36	Sequence 36, Appl
273	7	1.6	982	9	US-09-991-496-95	Sequence 95, Appl	346	6	1.4	8	9	US-10-057-321-2	Sequence 2, Appli
274	7	1.6	982	10	US-09-874-923-95	Sequence 95, Appl	347	6	1.4	8	9	US-09-978-917A-42	Sequence 42, Appl
275	7	1.6	1079	9	US-09-895-793-947	Sequence 947, App	348	6	1.4	8	9	US-09-967-237-22	Sequence 22, Appl
276	7	1.6	1079	10	US-09-822-827-947	Sequence 947, App	349	6	1.4	8	9	US-10-191-879-15	Sequence 15, Appl
277	7	1.6	1115	10	US-09-771-161A-253	Sequence 253, App	350	6	1.4	8	9	US-09-760-008A-10	Sequence 10, Appl
278	7	1.6	1427	9	US-09-991-496-97	Sequence 97, Appl	351	6	1.4	8	10	US-09-334-477-11	Sequence 11, Appl
279	7	1.6	1427	10	US-09-874-923-97	Sequence 97, Appl	352	6	1.4	8	10	US-09-244-984-2	Sequence 2, Appli
280	7	1.6	1641	9	US-09-991-496-96	Sequence 96, Appl	353	6	1.4	8	10	US-09-780-933-15	Sequence 15, Appl
281	7	1.6	1641	10	US-09-874-923-96	Sequence 96, Appl	354	6	1.4	8	10	US-09-772-719-22	Sequence 22, Appl
282	6	1.4	6	1	US-08-464-363-72	Sequence 72, Appl	355	6	1.4	8	12	US-10-003-496-10	Sequence 10, Appl
283	6	1.4	6	9	US-09-144-886-5	Sequence 5, Appli	356	6	1.4	9	9	US-10-171-417-8	Sequence 8, Appli
284	6	1.4	6	9	US-10-158-895-6	Sequence 6, Appli	357	6	1.4	9	10	US-09-284-663A-25	Sequence 25, Appl
285	6	1.4	6	9	US-09-858-332-9	Sequence 9, Appli	358	6	1.4	9	10	US-09-809-517A-5	Sequence 5, Appli
286	6	1.4	6	9	US-09-935-868-1	Sequence 1, Appli	359	6	1.4	9	10	US-09-854-280-18	Sequence 18, Appl
287	6	1.4	6	9	US-10-023-009-17	Sequence 17, Appl	360	6	1.4	9	10	US-09-821-883-15	Sequence 15, Appl
288	6	1.4	6	9	US-09-877-650-18	Sequence 18, Appl	361	6	1.4	10	1	US-09-854-208-18	Sequence 73, Appl
289	6	1.4	6	9	US-10-024-918-23	Sequence 23, Appl	362	6	1.4	10	9	US-08-464-363-73	Sequence 18, Appl
290	6	1.4	6	9	US-10-093-200A-10	Sequence 10, Appl	363	6	1.4	10	9	US-10-029-009-18	Sequence 24, Appl
291	6	1.4	6	9	US-10-061-395-68	Sequence 68, Appl	364	6	1.4	10	9	US-10-057-505-24	Sequence 2, Appli
292	6	1.4	6	9	US-10-116-273-35	Sequence 35, Appl	365	6	1.4	10	9	US-09-981-636-2	Sequence 2, Appli
293	6	1.4	6	9	US-09-814-604-9	Sequence 9, Appli	366	6	1.4	10	9	US-09-981-636-3	Sequence 3, Appli
294	6	1.4	6	9	US-10-147-087-6	Sequence 6, Appli	367	6	1.4	10	9	US-10-022-832-77	Sequence 77, Appl
295	6	1.4	6	9	US-10-166-232A-6	Sequence 6, Appli	368	6	1.4	10	9	US-09-911-569-101	Sequence 101, App
296	6	1.4	6	9	US-09-948-391A-43	Sequence 43, Appl	369	6	1.4	10	9	US-10-047-991-10	Sequence 10, Appl
297	6	1.4	6	9	US-09-978-917A-41	Sequence 41, Appl	370	6	1.4	10	9	US-10-174-410-6	Sequence 6, Appli
298	6	1.4	6	9	US-10-083-815-1	Sequence 1, Appli	371	6	1.4	10	9	US-10-174-410-11	Sequence 11, Appl
299	6	1.4	6	9	US-10-077-023-93	Sequence 93, Appl	372	6	1.4	10	10	US-09-809-517A-2	Sequence 2, Appli
300	6	1.4	6	9	US-10-015-979-12	Sequence 12, Appl	373	6	1.4	10	10	US-09-402-131-9	Sequence 9, Appli
301	6	1.4	6	9	US-10-071-838-13	Sequence 13, Appl	374	6	1.4	11	9	US-09-887-853-11	Sequence 11, Appl
302	6	1.4	6	9	US-09-989-981A-11	Sequence 11, Appl	375	6	1.4	11	9	US-10-105-717-4	Sequence 4, Appli
303	6	1.4	6	9	US-10-058-636-6	Sequence 6, Appli	376	6	1.4	11	10	US-09-814-569-2	Sequence 2, Appli
304	6	1.4	6	9	US-10-035-045-24	Sequence 24, Appl	377	6	1.4	12	9	US-09-927-436-1	Sequence 1, Appli
305	6	1.4	6	9	US-10-092-934-11	Sequence 11, Appl	378	6	1.4	12	9	US-10-171-417-4	Sequence 4, Appli
306	6	1.4	6	9	US-09-911-569-106	Sequence 106, Appl	379	6	1.4	12	10	US-09-946-893-3	Sequence 3, Appli
307	6	1.4	6	9	US-10-047-991-12	Sequence 12, Appl	380	6	1.4	13	10	US-09-821-883-17	Sequence 17, Appl
308	6	1.4	6	9	US-09-922-226-43	Sequence 43, Appl	381	6	1.4	13	10	US-09-402-131-10	Sequence 10, Appl
309	6	1.4	6	9	US-10-059-261-286	Sequence 261, App	382	6	1.4	13	10	US-09-823-266-2	Sequence 2, Appli
310	6	1.4	6	9	US-10-090-365-12	Sequence 12, Appl	383	6	1.4	13	10	US-09-823-266-3	Sequence 3, Appli
311	6	1.4	6	9	US-10-059-271-80	Sequence 80, Appl	384	6	1.4	13	10	US-09-823-266-4	Sequence 4, Appli

385	1.4	14	1	US-08-464-363-76	Sequence 76, Appl	458	6	1.4	27	9	US-10-226-956-283	Sequence 283, App
386	1.4	14	9	US-09-784-199-9	Sequence 9, Appl	459	6	1.4	27	9	US-10-226-956-289	Sequence 289, App
387	1.4	14	9	US-10-196-107A-58	Sequence 58, Appl	460	6	1.4	27	10	US-09-846-729A-22	Sequence 22, Appl
388	1.4	14	10	US-09-374-671-58	Sequence 58, Appl	461	6	1.4	27	10	US-09-846-729A-26	Sequence 26, Appl
389	1.4	14	10	US-09-823-266-1	Sequence 1, Appl	462	6	1.4	27	10	US-09-898-461-9	Sequence 9, Appl
390	1.4	14	10	US-09-900-530A-19	Sequence 19, Appl	463	6	1.4	28	9	US-10-092-750-152	Sequence 152, App
391	1.4	15	9	US-10-014-485A-43	Sequence 43, Appl	464	6	1.4	29	9	US-10-282-121-12	Sequence 12, Appl
392	1.4	15	9	US-10-010-184A-4	Sequence 4, Appl	465	6	1.4	29	10	US-09-880-901-2	Sequence 2, Appl
393	1.4	15	9	US-10-196-107A-53	Sequence 53, Appl	466	6	1.4	29	10	US-09-935-291A-24	Sequence 24, Appl
394	1.4	15	9	US-09-872-712-28	Sequence 2, Appl	467	6	1.4	30	9	US-09-911-569-98	Sequence 98, Appl
395	1.4	15	9	US-09-872-712-28	Sequence 28, Appl	468	6	1.4	30	9	US-09-911-569-100	Sequence 100, App
396	1.4	15	9	US-10-174-105A-43	Sequence 43, Appl	469	6	1.4	30	9	US-09-911-569-104	Sequence 104, App
397	1.4	15	10	US-09-374-671-53	Sequence 53, Appl	470	6	1.4	30	10	US-09-864-761-36251	Sequence 36251, A
398	1.4	15	10	US-09-821-883-16	Sequence 16, Appl	471	6	1.4	30	10	US-09-754-826-3	Sequence 3, Appl
399	1.4	15	12	US-10-139-841-43	Sequence 43, Appl	472	6	1.4	31	9	US-09-259-658-58	Sequence 58, Appl
400	1.4	16	9	US-09-911-569-103	Sequence 103, App	473	6	1.4	31	9	US-09-996-069-4	Sequence 4, Appl
401	1.4	16	9	US-09-911-569-105	Sequence 105, App	474	6	1.4	31	10	US-09-925-442-37	Sequence 37, Appl
402	1.4	16	10	US-09-809-517A-4	Sequence 4, Appl	475	6	1.4	32	9	US-10-174-410-182	Sequence 182, App
403	1.4	16	10	US-09-809-517A-29	Sequence 29, Appl	476	6	1.4	32	10	US-09-864-761-34706	Sequence 34706, A
404	1.4	16	10	US-09-351-794A-7	Sequence 7, Appl	477	6	1.4	33	9	US-09-259-658-49	Sequence 49, Appl
405	1.4	16	10	US-09-948-018-25	Sequence 25, Appl	478	6	1.4	33	9	US-09-259-658-50	Sequence 50, Appl
406	1.4	17	9	US-09-051-013-6	Sequence 6, Appl	479	6	1.4	33	9	US-09-866-538-21	Sequence 21, Appl
407	1.4	17	9	US-10-083-590-12	Sequence 12, Appl	480	6	1.4	33	10	US-10-121-258-22	Sequence 22, Appl
408	1.4	17	9	US-09-925-287-1	Sequence 1, Appl	481	6	1.4	33	10	US-09-864-761-36853	Sequence 36853, A
409	1.4	17	9	US-10-047-991-8	Sequence 8, Appl	482	6	1.4	35	9	US-10-083-815-68	Sequence 68, Appl
410	1.4	17	10	US-09-809-517A-32	Sequence 32, Appl	483	6	1.4	35	10	US-09-864-761-36426	Sequence 36426, A
411	1.4	18	1	US-08-954-771-47	Sequence 47, Appl	484	6	1.4	36	9	US-09-876-904A-309	Sequence 309, App
412	1.4	18	1	US-10-105-717-3	Sequence 3, Appl	485	6	1.4	36	9	US-09-276-455-8	Sequence 8, Appl
413	1.4	18	9	US-09-876-904A-92	Sequence 92, Appl	486	6	1.4	37	9	US-09-575-847-5	Sequence 5, Appl
414	1.4	18	10	US-09-809-517A-1	Sequence 1, Appl	487	6	1.4	37	9	US-09-911-569-97	Sequence 97, Appl
415	1.4	18	10	US-09-864-761-34373	Sequence 34373, A	488	6	1.4	37	10	US-09-864-761-40909	Sequence 40909, A
416	1.4	18	10	US-09-864-761-37659	Sequence 37659, A	489	6	1.4	39	9	US-09-996-069-1	Sequence 1, Appl
417	1.4	19	9	US-09-933-999A-31	Sequence 31, Appl	490	6	1.4	39	9	US-10-092-154-928	Sequence 928, App
418	1.4	19	9	US-09-996-069-13	Sequence 13, Appl	491	6	1.4	39	9	US-10-100-679-78	Sequence 78, Appl
419	1.4	19	9	US-09-876-904A-627	Sequence 627, App	492	6	1.4	39	10	US-10-102-806-429	Sequence 429, App
420	1.4	20	9	US-09-051-013-3	Sequence 3, Appl	493	6	1.4	39	10	US-09-884-681-39	Sequence 39, Appl
421	1.4	20	9	US-08-300-425B-26	Sequence 26, Appl	494	6	1.4	39	10	US-09-764-847-928	Sequence 928, App
422	1.4	21	8	US-08-326-626-13	Sequence 13, Appl	495	6	1.4	42	9	US-09-972-708-2	Sequence 2, Appl
423	1.4	21	9	US-10-082-659-16	Sequence 16, Appl	496	6	1.4	42	10	US-09-903-452-12	Sequence 12, Appl
424	1.4	21	9	US-09-092-296-20	Sequence 20, Appl	497	6	1.4	43	9	US-09-259-658-1	Sequence 1, Appl
425	1.4	21	9	US-09-104-408-37	Sequence 37, Appl	498	6	1.4	43	9	US-09-259-658-25	Sequence 25, Appl
426	1.4	21	9	US-10-216-408-27	Sequence 27, Appl	499	6	1.4	43	9	US-09-259-658-27	Sequence 27, Appl
427	1.4	21	9	US-09-939-126-8	Sequence 8, Appl	500	6	1.4	43	9	US-09-259-658-30	Sequence 30, Appl
428	1.4	21	9	US-09-939-126-9	Sequence 9, Appl	501	6	1.4	43	9	US-09-259-658-38	Sequence 38, Appl
429	1.4	21	9	US-09-939-126-10	Sequence 10, Appl	502	6	1.4	43	9	US-09-259-658-40	Sequence 40, Appl
430	1.4	21	9	US-09-939-126-11	Sequence 11, Appl	503	6	1.4	43	9	US-09-259-658-42	Sequence 42, Appl
431	1.4	21	9	US-09-939-126-12	Sequence 12, Appl	504	6	1.4	43	9	US-09-259-658-45	Sequence 45, Appl
432	1.4	21	9	US-10-278-547-49	Sequence 49, Appl	505	6	1.4	43	9	US-09-259-658-48	Sequence 48, Appl
433	1.4	21	10	US-09-050-516-49	Sequence 49, Appl	506	6	1.4	43	9	US-09-259-658-52	Sequence 52, Appl
434	1.4	21	10	US-09-276-600-11	Sequence 11, Appl	507	6	1.4	43	9	US-09-259-658-54	Sequence 54, Appl
435	1.4	21	10	US-09-099-823-27	Sequence 27, Appl	508	6	1.4	43	9	US-09-259-658-56	Sequence 56, Appl
436	1.4	21	10	US-09-234-717-25	Sequence 25, Appl	509	6	1.4	43	10	US-09-864-761-46454	Sequence 46454, A
437	1.4	21	10	US-09-850-178-19	Sequence 19, Appl	510	6	1.4	44	9	US-09-911-569-99	Sequence 99, Appl
438	1.4	21	10	US-09-193-538-23	Sequence 23, Appl	511	6	1.4	44	10	US-09-864-761-39281	Sequence 39281, A
439	1.4	21	10	US-09-250-883-23	Sequence 23, Appl	512	6	1.4	44	10	US-09-766-378A-34	Sequence 34, Appl
440	1.4	21	10	US-09-096-259-33	Sequence 33, Appl	513	6	1.4	45	9	US-09-921-144-6	Sequence 6, Appl
441	1.4	21	10	US-09-215-652-48	Sequence 48, Appl	514	6	1.4	45	10	US-09-766-378A-33	Sequence 33, Appl
442	1.4	21	10	US-09-347-064-18	Sequence 18, Appl	515	6	1.4	46	9	US-09-996-069-5	Sequence 5, Appl
443	1.4	21	10	US-09-049-695A-22	Sequence 22, Appl	516	6	1.4	46	9	US-09-764-891-4755	Sequence 4755, Ap
444	1.4	21	10	US-09-092-297-22	Sequence 22, Appl	517	6	1.4	47	9	US-09-801-089-18	Sequence 18, Appl
445	1.4	21	12	US-10-025-167-51	Sequence 51, Appl	518	6	1.4	47	9	US-10-173-570-5	Sequence 5, Appl
446	1.4	22	9	US-09-911-569-102	Sequence 102, App	519	6	1.4	47	10	US-09-738-945-5	Sequence 5, Appl
447	1.4	23	9	US-10-105-717-5	Sequence 5, Appl	520	6	1.4	47	10	US-09-923-243-1	Sequence 1, Appl
448	1.4	23	9	US-10-171-417-5	Sequence 5, Appl	521	6	1.4	48	9	US-09-921-144-8	Sequence 8, Appl
449	1.4	23	9	US-10-171-417-6	Sequence 6, Appl	522	6	1.4	48	10	US-09-864-761-37882	Sequence 37882, A
450	1.4	23	9	US-09-939-126-7	Sequence 7, Appl	523	6	1.4	51	9	US-09-896-720-22	Sequence 22, Appl
451	1.4	23	9	US-09-876-904A-624	Sequence 624, App	524	6	1.4	51	10	US-09-996-069-2	Sequence 2, Appl
452	1.4	24	9	US-09-876-904A-623	Sequence 623, App	525	6	1.4	52	10	US-09-864-761-38187	Sequence 38187, A
453	1.4	24	10	US-09-841-133-221	Sequence 221, App	526	6	1.4	52	10	US-09-864-761-42814	Sequence 42814, A
454	1.4	25	9	US-09-911-569-96	Sequence 96, Appl	527	6	1.4	54	9	US-09-986-480-209	Sequence 209, App
455	1.4	25	10	US-09-864-761-45218	Sequence 45218, A	528	6	1.4	54	9	US-09-996-069-3	Sequence 3, Appl
456	1.4	26	10	US-09-773-385-18	Sequence 18, Appl	529	6	1.4	54	10	US-09-925-302-817	Sequence 817, App
457	1.4	26	10	US-09-864-761-47310	Sequence 47310, A	530	6	1.4	54	10	US-09-864-761-38540	Sequence 38540, A

531	6	1.4	57	9	US-09-820-843A-3	Sequence 3, Appli	604	1.4	125	10	US-09-867-550-620	Sequence 620, App
532	6	1.4	58	10	US-09-864-761-37226	Sequence 37226, A	605	1.4	126	10	US-09-795-006A-43	Sequence 43, Appl
533	6	1.4	58	10	US-09-864-761-46792	Sequence 46792, A	606	1.4	126	10	US-09-795-006A-53	Sequence 53, Appl
534	6	1.4	58	10	US-09-864-761-48985	Sequence 48985, A	607	1.4	126	10	US-09-795-006A-55	Sequence 55, Appl
535	6	1.4	61	9	US-09-893-298-141	Sequence 141, App	608	1.4	126	10	US-09-795-006A-57	Sequence 57, Appl
536	6	1.4	64	10	US-09-925-297-652	Sequence 652, App	609	1.4	126	10	US-09-795-006A-61	Sequence 61, Appl
537	6	1.4	66	9	US-09-796-692-2334	Sequence 2334, Ap	610	1.4	126	10	US-09-795-006A-63	Sequence 63, Appl
538	6	1.4	66	9	US-10-040-862-2334	Sequence 2334, Ap	611	1.4	126	10	US-09-795-006A-65	Sequence 65, Appl
539	6	1.4	67	9	US-09-921-144-10	Sequence 10, Appl	612	1.4	127	10	US-09-795-006A-67	Sequence 67, Appl
540	6	1.4	67	10	US-09-988-899-8	Sequence 8, Appli	613	1.4	127	10	US-09-795-006A-83	Sequence 83, Appl
541	6	1.4	68	9	US-10-080-854-6	Sequence 6, Appli	614	1.4	127	10	US-09-795-006A-85	Sequence 85, Appl
542	6	1.4	70	9	US-09-921-144-12	Sequence 12, Appl	615	1.4	127	10	US-09-795-006A-87	Sequence 87, Appl
543	6	1.4	70	10	US-09-925-297-694	Sequence 694, App	616	1.4	127	10	US-09-795-006A-89	Sequence 89, Appl
544	6	1.4	71	12	US-10-001-870-139	Sequence 139, App	617	1.4	127	10	US-09-795-006A-91	Sequence 91, Appl
545	6	1.4	71	9	US-10-007-557-9	Sequence 9, Appli	618	1.4	127	10	US-09-795-006A-93	Sequence 93, Appl
546	6	1.4	72	9	US-09-820-843A-14	Sequence 14, Appl	619	1.4	127	10	US-09-795-006A-95	Sequence 95, Appl
547	6	1.4	74	10	US-09-864-761-35685	Sequence 35685, A	620	1.4	127	10	US-09-795-006A-97	Sequence 97, Appl
548	6	1.4	77	10	US-09-925-301-1464	Sequence 1464, Ap	621	1.4	128	10	US-09-742-373-9	Sequence 9, Appli
549	6	1.4	77	10	US-09-735-705-361	Sequence 361, App	622	1.4	128	10	US-09-795-006A-45	Sequence 45, Appl
550	6	1.4	77	10	US-09-850-716A-361	Sequence 361, App	623	1.4	128	10	US-09-795-006A-67	Sequence 67, Appl
551	6	1.4	77	10	US-09-897-778-361	Sequence 361, App	624	1.4	128	10	US-09-795-006A-69	Sequence 69, Appl
552	6	1.4	78	10	US-09-864-761-37352	Sequence 37352, A	625	1.4	128	10	US-09-795-006A-71	Sequence 71, Appl
553	6	1.4	80	9	US-10-080-854-4	Sequence 4, Appli	626	1.4	128	10	US-09-795-006A-73	Sequence 73, Appl
554	6	1.4	80	10	US-09-864-761-35659	Sequence 35659, A	627	1.4	128	10	US-09-795-006A-75	Sequence 75, Appl
555	6	1.4	81	9	US-10-006-593-53	Sequence 53, Appl	628	1.4	128	10	US-09-795-006A-79	Sequence 79, Appl
556	6	1.4	84	10	US-09-778-320-301	Sequence 301, App	629	1.4	129	10	US-09-911-781-5	Sequence 5, Appli
557	6	1.4	84	10	US-09-910-689-301	Sequence 301, App	630	1.4	129	10	US-09-795-006A-49	Sequence 49, Appl
558	6	1.4	84	12	US-10-010-742-301	Sequence 301, App	631	1.4	129	10	US-09-795-006A-99	Sequence 99, Appl
559	6	1.4	85	10	US-09-280-030-55	Sequence 65, Appl	632	1.4	129	10	US-09-795-006A-101	Sequence 101, App
560	6	1.4	87	10	US-09-864-761-33727	Sequence 33727, A	633	1.4	129	10	US-09-795-006A-103	Sequence 103, App
561	6	1.4	87	10	US-09-864-761-34744	Sequence 34744, A	634	1.4	129	10	US-09-795-006A-105	Sequence 105, App
562	6	1.4	89	9	US-09-764-891-5197	Sequence 5197, Ap	635	1.4	129	10	US-09-795-006A-107	Sequence 107, App
563	6	1.4	92	10	US-09-833-747A-2	Sequence 2, Appli	636	1.4	129	10	US-09-795-006A-109	Sequence 109, App
564	6	1.4	92	10	US-09-944-849-16	Sequence 16, Appl	637	1.4	129	10	US-09-795-006A-111	Sequence 111, App
565	6	1.4	93	9	US-10-125-540-325	Sequence 325, App	638	1.4	129	10	US-09-795-006A-113	Sequence 113, App
566	6	1.4	93	9	US-10-103-313-401	Sequence 401, App	639	1.4	130	10	US-09-280-030-62	Sequence 62, Appl
567	6	1.4	93	10	US-09-764-870-325	Sequence 325, App	640	1.4	131	9	US-09-982-992A-2	Sequence 2, Appli
568	6	1.4	93	10	US-09-764-853-617	Sequence 617, App	641	1.4	133	9	US-10-125-635A-344	Sequence 344, App
569	6	1.4	94	9	US-10-090-035-8	Sequence 8, Appli	642	1.4	133	9	US-09-938-864-344	Sequence 344, App
570	6	1.4	94	9	US-10-090-035-18	Sequence 18, Appl	643	1.4	133	9	US-09-785-019-344	Sequence 344, App
571	6	1.4	95	9	US-10-042-945-56	Sequence 56, Appl	644	1.4	133	9	US-09-949-780-4	Sequence 4, Appli
572	6	1.4	96	1	US-08-464-363-39	Sequence 39, Appl	645	1.4	135	9	US-09-738-626-5246	Sequence 5246, Ap
573	6	1.4	96	9	US-10-012-896-827	Sequence 827, App	646	1.4	138	10	US-09-881-752A-298	Sequence 298, App
574	6	1.4	96	9	US-09-895-793-827	Sequence 827, App	647	1.4	139	10	US-09-813-820-8	Sequence 8, Appli
575	6	1.4	96	9	US-09-895-814-827	Sequence 827, App	648	1.4	140	10	US-09-280-030-64	Sequence 64, Appl
576	6	1.4	96	10	US-09-759-143-827	Sequence 827, App	649	1.4	140	10	US-09-864-761-42549	Sequence 42549, A
577	6	1.4	96	10	US-09-780-669-827	Sequence 827, App	650	1.4	140	10	US-09-867-550-868	Sequence 868, App
578	6	1.4	96	10	US-09-822-827-827	Sequence 827, App	651	1.4	142	9	US-09-796-692-1189	Sequence 1189, Ap
579	6	1.4	97	9	US-10-080-854-2	Sequence 2, Appli	652	1.4	142	9	US-10-040-862-1189	Sequence 1189, Ap
580	6	1.4	97	10	US-09-334-477-23	Sequence 23, Appl	653	1.4	142	10	US-09-864-761-33485	Sequence 33485, A
581	6	1.4	97	10	US-09-334-477-27	Sequence 27, Appl	654	1.4	143	10	US-09-925-300-1867	Sequence 1867, Ap
582	6	1.4	98	10	US-09-850-564-4	Sequence 4, Appli	655	1.4	146	9	US-09-974-879-332	Sequence 332, App
583	6	1.4	98	10	US-10-125-635A-342	Sequence 342, App	656	1.4	146	10	US-09-923-304-9	Sequence 9, Appli
584	6	1.4	99	9	US-09-938-864-342	Sequence 342, App	657	1.4	148	9	US-09-828-523A-94	Sequence 94, Appl
585	6	1.4	99	9	US-09-783-019-342	Sequence 342, App	658	1.4	149	9	US-09-738-626-6297	Sequence 6297, Ap
586	6	1.4	99	10	US-09-864-761-36007	Sequence 36007, A	659	1.4	151	10	US-09-864-761-42051	Sequence 42051, A
587	6	1.4	101	9	US-10-042-945-58	Sequence 58, Appl	660	1.4	152	9	US-09-738-626-3689	Sequence 3689, Ap
588	6	1.4	102	9	US-10-042-945-59	Sequence 59, Appl	661	1.4	152	9	US-10-125-635A-343	Sequence 343, App
589	6	1.4	102	9	US-10-300-616-54	Sequence 54, Appl	662	1.4	152	9	US-09-938-864-343	Sequence 343, App
590	6	1.4	104	9	US-09-870-759-90	Sequence 90, Appl	663	1.4	152	9	US-09-785-019-343	Sequence 343, App
591	6	1.4	105	10	US-09-864-761-38933	Sequence 38933, A	664	1.4	152	10	US-09-927-738-22	Sequence 22, Appl
592	6	1.4	107	9	US-10-083-815-71	Sequence 71, Appl	665	1.4	155	9	US-10-121-988-181	Sequence 181, App
593	6	1.4	108	9	US-10-104-755-74	Sequence 74, Appl	666	1.4	156	9	US-10-043-487-239	Sequence 239, App
594	6	1.4	108	9	US-10-104-755-76	Sequence 76, Appl	667	1.4	156	10	US-09-864-761-41679	Sequence 41679, A
595	6	1.4	111	9	US-09-948-391A-9	Sequence 9, Appli	668	1.4	157	10	US-09-854-731-6	Sequence 6, Appli
596	6	1.4	114	10	US-09-864-761-35330	Sequence 35330, A	669	1.4	159	10	US-09-813-820-2	Sequence 2, Appli
597	6	1.4	114	10	US-09-864-761-36846	Sequence 36846, A	670	1.4	159	10	US-09-934-706-2	Sequence 2, Appli
598	6	1.4	117	9	US-09-948-391A-22	Sequence 22, Appl	671	1.4	161	9	US-10-012-896-846	Sequence 846, App
599	6	1.4	118	10	US-09-205-658-120	Sequence 120, App	672	1.4	161	9	US-09-895-793-846	Sequence 846, App
600	6	1.4	119	9	US-10-103-313-433	Sequence 433, App	673	1.4	161	9	US-09-895-814-846	Sequence 846, App
601	6	1.4	120	9	US-10-162-127-2	Sequence 2, Appli	674	1.4	161	9	US-10-101-464A-676	Sequence 676, App
602	6	1.4	120	9	US-10-293-705-2	Sequence 2, Appli	675	1.4	161	10	US-09-759-143-846	Sequence 846, App
603	6	1.4	121	9	US-10-162-127-4	Sequence 4, Appli	676	1.4	161	10	US-09-780-669-846	Sequence 846, App

677	1.4	161	10	US-09-925-297-658	Sequence 658, App	750	1.4	217	9	US-09-738-626-5827	Sequence 5827, App
678	1.4	161	10	US-09-822-827-846	Sequence 846, App	751	1.4	217	10	US-09-815-837-88	Sequence 88, Appli
679	1.4	161	10	US-09-850-716A-423	Sequence 423, App	752	1.4	220	9	US-10-000-903-1	Sequence 1, Appli
680	1.4	161	10	US-09-897-778-423	Sequence 423, App	753	1.4	220	9	US-10-000-903-8	Sequence 8, Appli
681	1.4	162	9	US-09-921-144-16	Sequence 16, Appli	754	1.4	221	9	US-10-169-048-2	Sequence 2, Appli
682	1.4	163	9	US-09-921-144-14	Sequence 14, Appli	755	1.4	222	9	US-10-149-819-1	Sequence 1, Appli
683	1.4	164	9	US-10-114-893-42	Sequence 42, Appli	756	1.4	223	9	US-09-738-626-4214	Sequence 19, Appli
684	1.4	166	9	US-09-966-546-8	Sequence 8, Appli	757	1.4	223	9	US-10-062-523-19	Sequence 19, Appli
685	1.4	166	9	US-09-966-546-30	Sequence 30, Appli	758	1.4	223	10	US-09-815-242-5788	Sequence 5788, Ap
686	1.4	166	9	US-09-966-545-8	Sequence 8, Appli	759	1.4	223	12	US-10-062-254-204	Sequence 204, App
687	1.4	166	9	US-09-966-545-30	Sequence 30, Appli	760	1.4	224	9	US-10-259-430-30	Sequence 30, Appli
688	1.4	166	9	US-10-012-896-838	Sequence 838, App	761	1.4	225	9	US-10-173-570-7	Sequence 7, Appli
689	1.4	166	9	US-09-835-793-838	Sequence 838, App	762	1.4	225	9	US-10-173-570-8	Sequence 8, Appli
690	1.4	166	9	US-09-895-814-838	Sequence 838, App	763	1.4	225	10	US-09-738-945-7	Sequence 7, Appli
691	1.4	166	9	US-09-965-212-8	Sequence 8, Appli	764	1.4	225	10	US-09-738-945-8	Sequence 8, Appli
692	1.4	166	9	US-09-965-212-30	Sequence 30, Appli	765	1.4	227	9	US-10-000-903-16	Sequence 16, Appli
693	1.4	166	10	US-09-731-221-77	Sequence 77, Appli	766	1.4	227	9	US-10-000-903-19	Sequence 19, Appli
694	1.4	166	10	US-09-759-143-838	Sequence 838, App	767	1.4	227	9	US-10-173-570-9	Sequence 9, Appli
695	1.4	166	10	US-09-780-669-838	Sequence 838, App	768	1.4	227	10	US-09-738-945-9	Sequence 9, Appli
696	1.4	166	10	US-09-822-827-838	Sequence 838, App	769	1.4	231	10	US-09-925-301-1336	Sequence 1336, Ap
697	1.4	167	9	US-09-902-941-1921	Sequence 1921, Ap	770	1.4	232	10	US-09-778-320-294	Sequence 294, App
698	1.4	167	9	US-09-849-626-1921	Sequence 1921, Ap	771	1.4	232	10	US-09-910-689-294	Sequence 294, App
699	1.4	167	9	US-10-017-754-1921	Sequence 1921, Ap	772	1.4	232	12	US-10-010-742-294	Sequence 294, App
700	1.4	168	9	US-09-862-540-48	Sequence 48, Appli	773	1.4	235	9	US-09-860-696-2	Sequence 2, Appli
701	1.4	168	10	US-09-815-242-13495	Sequence 13495, A	774	1.4	235	9	US-10-188-947-3	Sequence 3, Appli
702	1.4	170	9	US-10-174-410-12	Sequence 12, Appli	775	1.4	236	9	US-10-108-605-233	Sequence 233, App
703	1.4	171	9	US-10-156-275-76	Sequence 76, Appli	776	1.4	238	10	US-09-815-242-12790	Sequence 12790, A
704	1.4	173	9	US-09-738-626-6034	Sequence 6034, Ap	777	1.4	239	9	US-10-000-903-12	Sequence 12, Appli
705	1.4	175	10	US-09-854-280-23	Sequence 23, Appli	778	1.4	241	9	US-09-895-913A-274	Sequence 7, Appli
706	1.4	175	10	US-09-854-208-23	Sequence 23, Appli	779	1.4	241	9	US-09-895-913A-274	Sequence 274, App
707	1.4	177	9	US-10-082-830-204	Sequence 204, App	780	1.4	241	10	US-09-815-242-13469	Sequence 13469, A
708	1.4	178	9	US-09-813-718-6	Sequence 6, Appli	781	1.4	241	10	US-09-815-242-13469	Sequence 842, App
709	1.4	181	10	US-09-864-761-33314	Sequence 33314, A	782	1.4	241	9	US-09-895-793-842	Sequence 842, App
710	1.4	181	10	US-09-864-761-41658	Sequence 41658, A	783	1.4	241	9	US-09-895-814-842	Sequence 842, App
711	1.4	182	10	US-09-815-242-5853	Sequence 5853, Ap	784	1.4	241	10	US-09-759-143-842	Sequence 842, App
712	1.4	182	10	US-09-815-242-12718	Sequence 12718, A	785	1.4	241	10	US-09-780-669-842	Sequence 842, App
713	1.4	183	9	US-10-124-903-2	Sequence 2, Appli	786	1.4	241	10	US-09-822-827-842	Sequence 842, App
714	1.4	184	9	US-09-921-144-20	Sequence 20, Appli	787	1.4	243	9	US-09-828-523A-50	Sequence 50, Appli
715	1.4	184	9	US-09-345-373-30	Sequence 30, Appli	788	1.4	243	9	US-10-025-380-1122	Sequence 1122, Ap
716	1.4	185	9	US-09-895-913A-142	Sequence 142, App	789	1.4	243	9	US-09-900-345A-50	Sequence 50, Appli
717	1.4	185	9	US-09-921-144-18	Sequence 18, Appli	790	1.4	243	9	US-09-900-345A-52	Sequence 52, Appli
718	1.4	188	9	US-09-828-523A-58	Sequence 58, Appli	791	1.4	243	10	US-09-922-217-1122	Sequence 1122, Ap
719	1.4	189	9	US-09-895-913A-248	Sequence 248, App	792	1.4	243	10	US-09-924-099-10	Sequence 10, Appli
720	1.4	192	9	US-09-764-891-5030	Sequence 5030, Ap	793	1.4	243	9	US-10-267-311-23	Sequence 23, Appli
721	1.4	192	9	US-10-103-313-503	Sequence 503, App	794	1.4	246	9	US-09-738-626-5646	Sequence 5646, Ap
722	1.4	192	10	US-09-841-132-442	Sequence 442, App	795	1.4	247	9	US-10-135-965-3	Sequence 3, Appli
723	1.4	195	10	US-09-858-664A-11	Sequence 11, Appli	796	1.4	248	9	US-09-925-299-973	Sequence 973, App
724	1.4	196	9	US-09-989-920-224	Sequence 224, App	797	1.4	248	9	US-10-162-127-8	Sequence 8, Appli
725	1.4	198	10	US-09-925-300-1562	Sequence 1562, Ap	798	1.4	248	10	US-09-925-299-973	Sequence 973, App
726	1.4	201	10	US-09-814-122-69	Sequence 69, Appli	799	1.4	249	9	US-09-792-793A-84	Sequence 84, Appli
727	1.4	203	9	US-09-809-391-383	Sequence 383, App	800	1.4	250	9	US-10-162-127-6	Sequence 6, Appli
728	1.4	205	9	US-10-054-683-23	Sequence 23, Appli	801	1.4	253	10	US-09-925-297-499	Sequence 499, App
729	1.4	206	9	US-09-797-500-7	Sequence 7, Appli	802	1.4	254	9	US-10-260-960-7	Sequence 7, Appli
730	1.4	206	10	US-09-854-280-24	Sequence 24, Appli	803	1.4	255	10	US-09-840-787-7	Sequence 7, Appli
731	1.4	206	10	US-09-815-837-86	Sequence 86, Appli	804	1.4	256	9	US-10-016-283-30	Sequence 30, Appli
732	1.4	206	10	US-09-854-208-24	Sequence 24, Appli	805	1.4	256	9	US-10-125-635A-335	Sequence 335, App
733	1.4	207	9	US-09-736-457-1667	Sequence 1667, Ap	806	1.4	256	9	US-09-938-864-335	Sequence 335, App
734	1.4	207	9	US-09-902-941-1667	Sequence 1667, Ap	807	1.4	256	9	US-09-785-019-335	Sequence 1, Appli
735	1.4	207	9	US-09-902-941-1913	Sequence 1913, Ap	808	1.4	257	10	US-09-814-569-1	Sequence 33, Appli
736	1.4	207	9	US-09-849-626-1667	Sequence 1667, Ap	809	1.4	259	10	US-09-742-693-33	Sequence 33, Appli
737	1.4	207	9	US-09-849-626-1913	Sequence 1913, Ap	810	1.4	259	10	US-09-925-301-1352	Sequence 1352, Ap
738	1.4	207	9	US-10-017-754-1667	Sequence 1667, Ap	811	1.4	260	9	US-09-764-868-1030	Sequence 1030, Ap
739	1.4	207	9	US-10-017-754-1913	Sequence 1913, Ap	812	1.4	260	10	US-09-742-693-32	Sequence 32, Appli
740	1.4	208	10	US-09-864-761-36456	Sequence 36456, A	813	1.4	260	10	US-09-742-693-32	Sequence 32, Appli
741	1.4	210	9	US-09-973-025-28	Sequence 28, Appli	814	1.4	261	10	US-09-940-037A-29	Sequence 29, Appli
742	1.4	211	9	US-09-899-303-28	Sequence 28, Appli	815	1.4	262	9	US-09-925-299-794	Sequence 794, App
743	1.4	211	9	US-10-193-460A-2	Sequence 2, Appli	816	1.4	262	10	US-09-925-299-794	Sequence 794, App
744	1.4	211	10	US-09-813-820-4	Sequence 4, Appli	817	1.4	262	10	US-09-895-828-457	Sequence 457, App
745	1.4	212	9	US-09-791-932-84	Sequence 84, Appli	818	1.4	265	10	US-09-742-693-30	Sequence 30, Appli
746	1.4	212	9	US-10-043-487-238	Sequence 238, App	819	1.4	265	10	US-09-732-091-44	Sequence 44, Appli
747	1.4	214	9	US-09-963-156A-1	Sequence 1, Appli	820	1.4	266	9	US-09-828-523A-78	Sequence 78, Appli
748	1.4	215	10	US-09-749-728B-23	Sequence 23, Appli	821	1.4	267	9	US-10-173-570-2	Sequence 2, Appli
749	1.4	216	9	US-10-016-283-31	Sequence 31, Appli	822	1.4	267	10	US-09-738-945-2	Sequence 2, Appli

823	6	1.4	269	10	US-09-888-721-36	Sequence 36, Appl	896	6	1.4	324	9	US-09-987-107-62	Sequence 62, Appl
824	6	1.4	271	10	US-10-156-932-9	Sequence 9, Appli	897	6	1.4	324	9	US-09-987-107-64	Sequence 64, Appl
825	6	1.4	271	10	US-09-864-761-37986	Sequence 37986, A	898	6	1.4	324	9	US-09-987-107-66	Sequence 66, Appl
826	6	1.4	273	9	US-09-987-107-50	Sequence 50, Appl	899	6	1.4	324	9	US-10-076-622-551	Sequence 551, App
827	6	1.4	273	9	US-10-000-903-4	Sequence 4, Appli	900	6	1.4	324	9	US-10-267-311-25	Sequence 25, Appl
828	6	1.4	273	10	US-09-816-714-4	Sequence 4, Appli	901	6	1.4	324	9	US-10-260-960-14	Sequence 14, Appl
829	6	1.4	278	9	US-10-000-903-21	Sequence 21, Appl	902	6	1.4	324	10	US-09-864-761-47729	Sequence 47729, A
830	6	1.4	278	9	US-10-150-469-4	Sequence 4, Appli	903	6	1.4	324	12	US-10-007-805-551	Sequence 551, App
831	6	1.4	279	9	US-10-121-988-90	Sequence 90, Appl	904	6	1.4	325	9	US-09-792-793A-81	Sequence 81, Appl
832	6	1.4	279	10	US-09-846-808-21	Sequence 21, Appl	905	6	1.4	326	9	US-09-991-211-1	Sequence 11, Appli
833	6	1.4	280	9	US-09-940-7278-119	Sequence 119, Appl	906	6	1.4	326	9	US-09-966-459A-14	Sequence 14, Appl
834	6	1.4	282	10	US-09-888-721-38	Sequence 38, Appl	907	6	1.4	326	10	US-09-334-477-25	Sequence 25, Appl
835	6	1.4	284	9	US-09-828-523A-54	Sequence 54, Appl	908	6	1.4	327	9	US-10-040-803-8	Sequence 8, Appli
836	6	1.4	285	9	US-10-278-173-16	Sequence 16, Appl	909	6	1.4	327	9	US-09-792-793A-72	Sequence 72, Appl
837	6	1.4	285	10	US-09-883-758-4	Sequence 4, Appli	910	6	1.4	327	9	US-09-792-793A-75	Sequence 75, Appl
838	6	1.4	286	9	US-10-195-142-12	Sequence 12, Appl	911	6	1.4	328	10	US-09-854-280-22	Sequence 22, Appl
839	6	1.4	286	10	US-09-862-915-1	Sequence 1, Appli	912	6	1.4	328	10	US-09-854-280-22	Sequence 22, Appl
840	6	1.4	287	10	US-09-888-721-40	Sequence 40, Appl	913	6	1.4	328	10	US-09-900-527-2	Sequence 2, Appli
841	6	1.4	288	10	US-09-895-919A-42	Sequence 42, Appl	914	6	1.4	329	10	US-09-802-674-13	Sequence 13, Appl
842	6	1.4	288	10	US-09-818-247-22	Sequence 22, Appl	915	6	1.4	331	9	US-09-828-523A-66	Sequence 66, Appl
843	6	1.4	290	10	US-09-808-885-1	Sequence 1, Appli	916	6	1.4	331	9	US-09-828-523A-70	Sequence 70, Appl
844	6	1.4	291	10	US-09-888-721-44	Sequence 44, Appl	917	6	1.4	331	9	US-09-824-438-6	Sequence 6, Appli
845	6	1.4	292	9	US-10-000-903-10	Sequence 10, Appl	918	6	1.4	331	9	US-10-059-261-169	Sequence 169, App
846	6	1.4	292	10	US-09-864-761-37944	Sequence 37944, A	919	6	1.4	333	9	US-10-059-261-61	Sequence 61, Appl
847	6	1.4	294	9	US-10-151-832-7	Sequence 7, Appli	920	6	1.4	334	9	US-10-017-736-4	Sequence 4, Appli
848	6	1.4	294	10	US-09-767-037-3	Sequence 3, Appli	921	6	1.4	334	9	US-10-142-356-12	Sequence 12, Appl
849	6	1.4	295	9	US-10-125-540-364	Sequence 364, App	922	6	1.4	335	9	US-09-850-258-2	Sequence 2, Appli
850	6	1.4	295	10	US-09-764-870-364	Sequence 364, App	923	6	1.4	335	10	US-09-850-260-2	Sequence 2, Appli
851	6	1.4	296	10	US-09-815-242-4892	Sequence 4892, Ap	924	6	1.4	335	10	US-09-850-260-2	Sequence 2, Appli
852	6	1.4	296	10	US-09-815-242-10483	Sequence 10483, A	925	6	1.4	335	10	US-09-850-260-2	Sequence 2, Appli
853	6	1.4	296	10	US-09-888-721-42	Sequence 42, Appl	926	6	1.4	335	10	US-09-886-055-345	Sequence 345, App
854	6	1.4	298	9	US-09-910-600-26	Sequence 26, Appl	927	6	1.4	336	9	US-10-012-896-940	Sequence 940, App
855	6	1.4	298	10	US-09-823-758-2	Sequence 2, Appli	928	6	1.4	336	9	US-09-895-793-940	Sequence 940, App
856	6	1.4	298	10	US-09-925-300-1486	Sequence 1486, Ap	929	6	1.4	336	9	US-09-895-793-940	Sequence 940, App
857	6	1.4	299	10	US-09-886-055-257	Sequence 257, App	930	6	1.4	336	10	US-09-780-669-940	Sequence 940, App
858	6	1.4	300	10	US-09-954-697-36	Sequence 36, Appl	931	6	1.4	336	10	US-09-822-827-940	Sequence 940, App
859	6	1.4	304	9	US-09-898-586-44	Sequence 44, Appl	932	6	1.4	337	9	US-09-813-153-241	Sequence 241, App
860	6	1.4	304	10	US-09-761-288-44	Sequence 44, Appl	933	6	1.4	338	9	US-10-027-805-40	Sequence 40, Appl
861	6	1.4	306	10	US-09-815-242-11035	Sequence 11035, A	934	6	1.4	338	9	US-10-027-804-40	Sequence 40, Appl
862	6	1.4	308	9	US-10-121-988-97	Sequence 97, Appl	935	6	1.4	338	10	US-09-903-410-40	Sequence 40, Appl
863	6	1.4	309	12	US-10-052-798-9	Sequence 9, Appli	936	6	1.4	339	9	US-10-195-142-4	Sequence 4, Appli
864	6	1.4	310	9	US-10-151-832-3	Sequence 3, Appli	937	6	1.4	341	9	US-10-176-664-3	Sequence 3, Appli
865	6	1.4	310	10	US-09-815-242-11064	Sequence 11064, A	938	6	1.4	341	9	US-10-176-664-6	Sequence 6, Appli
866	6	1.4	310	12	US-10-052-798-11	Sequence 11, Appl	939	6	1.4	341	9	US-10-176-664-6	Sequence 6, Appli
867	6	1.4	311	9	US-09-828-523A-86	Sequence 86, Appl	940	6	1.4	341	9	US-10-176-664-9	Sequence 9, Appli
868	6	1.4	311	9	US-09-893-519A-54	Sequence 54, Appl	941	6	1.4	341	9	US-10-176-664-11	Sequence 11, Appl
869	6	1.4	312	9	US-09-738-626-5731	Sequence 5731, Ap	942	6	1.4	341	10	US-09-401-636-3	Sequence 3, Appli
870	6	1.4	312	9	US-10-151-832-9	Sequence 9, Appli	943	6	1.4	341	10	US-09-401-636-4	Sequence 4, Appli
871	6	1.4	312	12	US-10-052-798-10	Sequence 10, Appl	944	6	1.4	341	10	US-09-401-636-6	Sequence 6, Appli
872	6	1.4	314	9	US-09-738-626-6809	Sequence 6809, Ap	945	6	1.4	341	10	US-09-401-636-9	Sequence 9, Appli
873	6	1.4	315	9	US-10-041-006A-9	Sequence 9, Appli	946	6	1.4	341	10	US-09-401-636-11	Sequence 11, Appl
874	6	1.4	315	9	US-09-940-925A-91	Sequence 91, Appl	947	6	1.4	342	9	US-10-176-664-5	Sequence 5, Appli
875	6	1.4	315	10	US-09-886-055-193	Sequence 193, App	948	6	1.4	342	9	US-10-176-664-8	Sequence 8, Appli
876	6	1.4	315	12	US-10-040-655-9	Sequence 9, Appli	949	6	1.4	342	10	US-09-401-636-5	Sequence 5, Appli
877	6	1.4	316	9	US-09-987-107-48	Sequence 48, Appl	950	6	1.4	342	10	US-09-401-636-8	Sequence 8, Appli
878	6	1.4	316	9	US-09-987-107-54	Sequence 54, Appl	951	6	1.4	343	9	US-10-176-664-7	Sequence 7, Appli
879	6	1.4	316	12	US-10-041-400A-9	Sequence 9, Appli	952	6	1.4	343	9	US-10-159-151-8	Sequence 8, Appli
880	6	1.4	316	12	US-10-041-264A-9	Sequence 9, Appli	953	6	1.4	343	10	US-09-401-636-7	Sequence 7, Appli
881	6	1.4	316	12	US-10-042-091A-9	Sequence 9, Appli	954	6	1.4	343	9	US-09-971-798-27	Sequence 27, Appl
882	6	1.4	317	9	US-10-151-832-5	Sequence 5, Appli	955	6	1.4	344	9	US-09-987-107-68	Sequence 68, Appl
883	6	1.4	318	10	US-09-759-281B-1	Sequence 1, Appli	956	6	1.4	345	9	US-10-176-664-10	Sequence 10, Appl
884	6	1.4	320	9	US-10-033-297-61	Sequence 61, Appl	957	6	1.4	345	10	US-09-401-636-10	Sequence 10, Appl
885	6	1.4	320	9	US-09-940-244-61	Sequence 61, Appl	958	6	1.4	345	10	US-09-813-820-7	Sequence 7, Appli
886	6	1.4	322	9	US-09-792-793A-78	Sequence 78, Appl	959	6	1.4	348	9	US-09-910-600-22	Sequence 22, Appl
887	6	1.4	323	9	US-09-987-107-56	Sequence 56, Appl	960	6	1.4	348	9	US-09-910-600-23	Sequence 23, Appl
888	6	1.4	323	9	US-09-987-107-58	Sequence 58, Appl	961	6	1.4	348	9	US-09-910-600-24	Sequence 24, Appl
889	6	1.4	323	9	US-09-987-107-60	Sequence 60, Appl	962	6	1.4	348	9	US-09-910-600-25	Sequence 25, Appl
890	6	1.4	323	9	US-10-076-622-483	Sequence 483, App	963	6	1.4	353	10	US-09-919-172-69	Sequence 69, Appl
891	6	1.4	323	10	US-09-813-153-132	Sequence 132, App	964	6	1.4	354	9	US-09-892-877-120	Sequence 120, App
892	6	1.4	323	10	US-09-334-477-21	Sequence 21, Appl	965	6	1.4	354	10	US-09-823-153-8	Sequence 8, Appli
893	6	1.4	323	10	US-09-604-287A-483	Sequence 483, App	966	6	1.4	354	10	US-09-925-302-501	Sequence 501, App
894	6	1.4	323	10	US-09-971-798-31	Sequence 31, Appl	967	6	1.4	357	9	US-10-012-896-826	Sequence 826, App
895	6	1.4	323	12	US-10-007-805-483	Sequence 483, App	968	6	1.4	357	9	US-09-895-793-826	Sequence 826, App

969 6 1.4 357 9 US-09-895-814-826 Sequence 826, App
970 6 1.4 357 10 US-09-759-143-826 Sequence 826, App
971 6 1.4 357 10 US-09-780-669-826 Sequence 826, App
972 6 1.4 357 10 US-09-822-827-826 Sequence 826, App
973 6 1.4 358 10 US-09-801-368-148 Sequence 148, App
974 6 1.4 359 10 US-09-764-864-996 Sequence 996, App
975 6 1.4 361 9 US-09-839-894-10 Sequence 10, Appl
976 6 1.4 361 9 US-09-839-894-28 Sequence 28, Appl
977 6 1.4 363 9 US-10-159-151-10 Sequence 10, Appl
978 6 1.4 364 9 US-09-829-275-1 Sequence 1, Appl
979 6 1.4 366 10 US-09-887-569A-2 Sequence 2, Appl
980 6 1.4 368 9 US-10-084-700-30 Sequence 30, Appl
981 6 1.4 370 9 US-09-828-523A-74 Sequence 74, Appl
982 6 1.4 371 9 US-10-000-903-6 Sequence 6, Appl
983 6 1.4 371 10 US-09-850-716A-433 Sequence 433, App
984 6 1.4 371 10 US-09-897-778-433 Sequence 433, App
985 6 1.4 372 9 US-09-813-718-4 Sequence 4, Appl
986 6 1.4 372 9 US-09-813-718-8 Sequence 8, Appl
987 6 1.4 373 9 US-10-159-151-4 Sequence 11, Appl
988 6 1.4 375 10 US-09-764-864-1316 Sequence 941, App
989 6 1.4 378 1 US-08-979-847-122 Sequence 1316, App
990 6 1.4 378 10 US-09-764-864-863 Sequence 863, App
991 6 1.4 379 10 US-09-134-333-11 Sequence 11, Appl
992 6 1.4 381 9 US-10-012-896-941 Sequence 941, App
993 6 1.4 381 9 US-09-895-793-941 Sequence 941, App
994 6 1.4 381 9 US-09-895-814-941 Sequence 941, App
995 6 1.4 381 9 US-10-159-151-6 Sequence 6, Appl
996 6 1.4 381 10 US-09-780-669-941 Sequence 941, App
997 6 1.4 381 10 US-09-822-827-941 Sequence 941, App
998 6 1.4 383 9 US-10-000-903-23 Sequence 23, Appl
999 6 1.4 383 10 US-09-883-096-5 Sequence 5, Appl
1000 6 1.4 387 9 US-09-972-268-16 Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-978-756-3
; Sequence 3, Application US/09978756
; Patent No. US20020160017A1
; GENERAL INFORMATION:
; APPLICANT: Holder, Anthony
; APPLICANT: Birdsell, Berry
; APPLICANT: Feeney, James
; APPLICANT: Morgan, William
; APPLICANT: Syed, Shabih
; TITLE OF INVENTION: Malaria Vaccine
; FILE REFERENCE: 18396/1005
; CURRENT APPLICATION NUMBER: US/09/978,756
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: PCT/GB00/01558
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 09/311,817
; PRIOR FILING DATE: 1999-05-13
; PRIOR APPLICATION NUMBER: 2,271,451
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 9909072.2
; PRIOR FILING DATE: 1999-04-20
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-978-756-3

Query Match 86.8%; Score 374; DB 9; Length 394;
Best Local Similarity 100.0%; Pred. No. 3.9e-293;
Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISVTMDNILSGFENEYDVYILKPLAGVYRSILKKQIEKNIFTNLDILNSRLKRRKY 117
|||||

Db 1 AISVTMDNILSGFENEYDVYILKPLAGVYRSILKKQIEKNIFTNLDILNSRLKRRKY 60
QY 118 FLDVLESDLMQFKHISNEYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKKFAQEGIS 177
|||||
Db 61 FLDVLESDLMQFKHISNEYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKKFAQEGIS 120
QY 178 YEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNITL 237
|||||
Db 121 YEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNITL 180
QY 238 YNNLVNKIDDDYILNKLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDEAIKK 297
|||||
Db 181 YNNLVNKIDDDYILNKLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNPYDEAIKK 240
QY 298 LINDTTRKMDLGLKLLSTGLVQNFNTIISKLTIEGKFQDMLNISQHCYKVKQCPNSGCFR 357
|||||
Db 241 LINDTTRKMDLGLKLLSTGLVQNFNTIISKLTIEGKFQDMLNISQHCYKVKQCPNSGCFR 300
QY 358 HLDRECECKLLNYKQEGDKCVENPPTCENNNGGCDADATCTEEDSGSSRKKITCECTK 417
|||||
Db 301 HLDRECECKLLNYKQEGDKCVENPPTCENNNGGCDADATCTEEDSGSSRKKITCECTK 360
QY 418 PDSYPLFDGIFCSS 431
|||||
Db 361 PDSYPLFDGIFCSS 374
RESULT 2
US-10-098-514-14
; Sequence 14, Application US/10098514
; Publication No. US20020194648A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra P
; APPLICANT: Christopher, David A
; APPLICANT: Vine, Benjamin
; APPLICANT: Su, Wei-Wen
; APPLICANT: Bugos, Robert
; TITLE OF INVENTION: PLASMODIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PROD
; FILE REFERENCE: A-71339/RFT/TAL/NBC
; CURRENT APPLICATION NUMBER: US/10/098,514
; CURRENT FILING DATE: 2002-08-06
; PRIOR APPLICATION NUMBER: US 09/500,376
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: US 60/274,599
; PRIOR FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
US-10-098-514-14

Query Match 53.6%; Score 231; DB 9; Length 402;
Best Local Similarity 100.0%; Pred. No. 4.3e-178;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISVTMDNILSGFENEYDVYILKPLAGVYRSILKKQIEKNIFTNLDILNSRLKRRKY 117
|||||
Db 22 AISVTMDNILSGFENEYDVYILKPLAGVYRSILKKQIEKNIFTNLDILNSRLKRRKY 81
QY 118 FLDVLESDLMQFKHISNEYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKKFAQEGIS 177
|||||
Db 82 FLDVLESDLMQFKHISNEYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKKFAQEGIS 141
QY 178 YEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNITL 237
|||||
Db 142 YEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNITL 201
QY 238 YNNLVNKIDDDYILNKLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKN 288

```
Db 202 YNNLVNKIDDYLNKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKN 252
|||||
RESULT 3
US-10-098-514-4
; Sequence 4, Application US/10098514
; Publication No. US20020194648A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra P
; APPLICANT: Christopher, David A
; APPLICANT: Vine, Benjamin
; APPLICANT: Su, Wei-Wen
; APPLICANT: Bugos, Robert
; TITLE OF INVENTION: PLASMODIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PRODUCE
; FILE REFERENCE: A-71339/RET/TAL/NEC
; CURRENT APPLICATION NUMBER: US/10/098,514
; CURRENT FILING DATE: 2002-08-06
; PRIOR APPLICATION NUMBER: US 09/500,376
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: US 60/274,599
; PRIOR FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
; NAME/KEY: MISC_FEATURE
; LOCATION: (380)..(380)
; OTHER INFORMATION: "xaa" at position 380 represents a stop codon
US-10-098-514-4
Query Match 52.7%; Score 227; DB 9; Length 383;
Best Local Similarity 100.0%; Pred. No. 6.9e-175;
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 TMDNLSGFENEYDVYILKPLAGVYRSLLKQIEKNITFNLDILNSRLKRRKYFLDV 121
|||||
Db 7 TMDNLSGFENEYDVYILKPLAGVYRSLLKQIEKNITFNLDILNSRLKRRKYFLDV 66
|||||
QY 122 LESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKFAQEGISYIEK 181
|||||
Db 67 LESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKFAQEGISYIEK 126
|||||
QY 182 VLAKYKDDLESIKKVIKEEKEFPSPPTPPSPAKTDEQKESKFLPFLTNIETLYNNL 241
|||||
Db 127 VLAKYKDDLESIKKVIKEEKEFPSPPTPPSPAKTDEQKESKFLPFLTNIETLYNNL 186
|||||
QY 242 VNKIDDYLNKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKN 288
|||||
Db 187 VNKIDDYLNKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKN 233
|||||
RESULT 4
US-10-098-514-2
; Sequence 2, Application US/10098514
; Publication No. US20020194648A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra P
; APPLICANT: Christopher, David A
; APPLICANT: Vine, Benjamin
; APPLICANT: Su, Wei-Wen
; APPLICANT: Bugos, Robert
; TITLE OF INVENTION: PLASMODIUM FALCIPARUM MEROZOITE SURFACE PROTEIN-1 MALARIA PRODUCE
; FILE REFERENCE: A-71339/RET/TAL/NEC
; CURRENT APPLICATION NUMBER: US/10/098,514
; CURRENT FILING DATE: 2002-08-06
; QUERY MATCH
Query Match 13.9%; Score 60; DB 10; Length 95;
Best Local Similarity 100.0%; Pred. No. 4.8e-41;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
; PRIOR APPLICATION NUMBER: US 09/500,376
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: US 60/274,599
; PRIOR FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
; NAME/KEY: MISC_FEATURE
; LOCATION: (380)..(380)
; OTHER INFORMATION: "xaa" at position 380 represents a stop codon
US-10-098-514-2
Query Match 52.2%; Score 225; DB 9; Length 383;
Best Local Similarity 100.0%; Pred. No. 2.8e-173;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 DNILSGFENEYDVYILKPLAGVYRSLLKQIEKNITFNLDILNSRLKRRKYFLDVLE 123
|||||
Db 9 DNILSGFENEYDVYILKPLAGVYRSLLKQIEKNITFNLDILNSRLKRRKYFLDVLE 68
|||||
QY 124 SLDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKFAQEGISYIEKVL 183
|||||
Db 69 SLDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSKYIKESVENDIKFAQEGISYIEKVL 128
|||||
QY 184 AKYKDDLESIKKVIKEEKEFPSPPTPPSPAKTDSQKESKFLPFLTNIETLYNNLVN 243
|||||
Db 129 AKYKDDLESIKKVIKEEKEFPSPPTPPSPAKTDSQKESKFLPFLTNIETLYNNLVN 188
|||||
QY 244 KIDDYLNKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKN 288
|||||
Db 189 KIDDYLNKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKN 233
|||||
RESULT 5
US-09-134-333-2
; Sequence 2, Application US/09134333
; Patent No. US20020076403A1
; GENERAL INFORMATION:
; APPLICANT: LONGACRE-ANDRE, SHIRLEY
; APPLICANT: ROTH, CHARLES
; APPLICANT: NATO, FARIDABANO
; APPLICANT: BARNWELL, JOHN
; APPLICANT: MENDIS, KAMINI
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF
; FILE REFERENCE: 0660-0135-0XCIP
; CURRENT APPLICATION NUMBER: US/09/134,333
; CURRENT FILING DATE: 1999-04-18
; EARLIER APPLICATION NUMBER: PCT/FR97/00290
; EARLIER FILING DATE: 1997-02-14
; EARLIER APPLICATION NUMBER: FR96/01822
; EARLIER FILING DATE: 1996-02-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-134-333-2
Query Match 13.9%; Score 60; DB 10; Length 95;
Best Local Similarity 100.0%; Pred. No. 4.8e-41;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 NISQHCQVKQCPENSGCFRHLDERECKLLNYKQESDKCVENPNPTCNENGGCDADA 397
```

Db 3 NISQHCYKQCPENSGCFRHLDERECKLLNLYKQEGDKCVENPNTCENNNGCDADA 62
|||||

RESULT 6
US-09-134-333-10
; Sequence 10, Application US/09134333
; Patent No. US20020076403A1
; GENERAL INFORMATION:
; APPLICANT: LONGACRE-ANDRE, SHIRLEY
; APPLICANT: ROTH, CHARLES
; APPLICANT: NATO, FARIDABANO
; APPLICANT: BARNWELL, JOHN
; APPLICANT: MENDIS, KAMINI
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF
; FILE REFERENCE: 0660-0135-0XCIP
; CURRENT APPLICATION NUMBER: US/09/134,333
; CURRENT FILING DATE: 1999-04-18
; EARLIER APPLICATION NUMBER: PCT/FR97/00290
; EARLIER FILING DATE: 1997-02-14
; EARLIER APPLICATION NUMBER: FR96/01822
; EARLIER FILING DATE: 1996-02-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-134-333-10

Query Match 13.9%; Score 60; DB 10; Length 108;
Best Local Similarity 100.0%; Pred. No. 5.3e-41;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 338 NISQHCYKQCPENSGCFRHLDERECKLLNLYKQEGDKCVENPNTCENNNGCDADA 397
|||||

Db 16 NISQHCYKQCPENSGCFRHLDERECKLLNLYKQEGDKCVENPNTCENNNGCDADA 75
|||||

RESULT 7
US-09-134-333-5
; Sequence 5, Application US/09134333
; Patent No. US20020076403A1
; GENERAL INFORMATION:
; APPLICANT: LONGACRE-ANDRE, SHIRLEY
; APPLICANT: ROTH, CHARLES
; APPLICANT: NATO, FARIDABANO
; APPLICANT: BARNWELL, JOHN
; APPLICANT: MENDIS, KAMINI
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF
; FILE REFERENCE: 0660-0135-0XCIP
; CURRENT APPLICATION NUMBER: US/09/134,333
; CURRENT FILING DATE: 1999-04-18
; EARLIER APPLICATION NUMBER: PCT/FR97/00290
; EARLIER FILING DATE: 1997-02-14
; EARLIER APPLICATION NUMBER: FR96/01822
; EARLIER FILING DATE: 1996-02-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 116
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-134-333-5

Query Match 13.9%; Score 60; DB 10; Length 116;
Best Local Similarity 100.0%; Pred. No. 5.6e-41;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 338 NISQHCYKQCPENSGCFRHLDERECKLLNLYKQEGDKCVENPNTCENNNGCDADA 397
|||||

Db 3 NISQHCYKQCPENSGCFRHLDERECKLLNLYKQEGDKCVENPNTCENNNGCDADA 62
|||||

RESULT 8
US-09-134-333-8
; Sequence 8, Application US/09134333
; Patent No. US20020076403A1
; GENERAL INFORMATION:
; APPLICANT: LONGACRE-ANDRE, SHIRLEY
; APPLICANT: ROTH, CHARLES
; APPLICANT: NATO, FARIDABANO
; APPLICANT: BARNWELL, JOHN
; APPLICANT: MENDIS, KAMINI
; TITLE OF INVENTION: RECOMBINANT PROTEIN CONTAINING A C-TERMINAL FRAGMENT OF
; FILE REFERENCE: 0660-0135-0XCIP
; CURRENT APPLICATION NUMBER: US/09/134,333
; CURRENT FILING DATE: 1999-04-18
; EARLIER APPLICATION NUMBER: PCT/FR97/00290
; EARLIER FILING DATE: 1997-02-14
; EARLIER APPLICATION NUMBER: FR96/01822
; EARLIER FILING DATE: 1996-02-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-134-333-8

Query Match 13.9%; Score 60; DB 10; Length 127;
Best Local Similarity 100.0%; Pred. No. 6.1e-41;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 338 NISQHCYKQCPENSGCFRHLDERECKLLNLYKQEGDKCVENPNTCENNNGCDADA 397
|||||

Db 35 NISQHCYKQCPENSGCFRHLDERECKLLNLYKQEGDKCVENPNTCENNNGCDADA 94
|||||

RESULT 9
US-09-804-626-4
; Sequence 4, Application US/09804626
; Patent No. US20020128190A1
; GENERAL INFORMATION:
; APPLICANT: Lobel, Leslie
; APPLICANT: Lustbader, Joyce
; TITLE OF INVENTION: EXPRESSION OF PROPERLY FOLDED AND SOLUBLE EXTRACELLULAR DOMAIN
; FILE REFERENCE: 0575/62259/JPW/SHS
; CURRENT APPLICATION NUMBER: US/09/804,626
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 516
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-804-626-4

Query Match 11.6%; Score 50; DB 10; Length 516;
Best Local Similarity 100.0%; Pred. No. 2.2e-32;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHISSGLVPRGSGMKETAAAKFERQHMDSPDLGTDGDDDDKAMADIGS 50
|||||

Db 116 MHHHHISSGLVPRGSGMKETAAAKFERQHMDSPDLGTDGDDDDKAMADIGS 165
|||||

RESULT 10
US-09-804-626-2
; Sequence 2, Application US/09804626

```
; Patent No. US20020128190A1
; GENERAL INFORMATION:
; APPLICANT: Lobel, Leslie
; APPLICANT: Lustbader, Joyce
; TITLE OF INVENTION: EXPRESSION OF PROPERLY FOLDED AND SOLUBLE EXTRACELLULAR DOMAIN OF
; TITLE OF INVENTION: CONADOTROPIN RECEPTOR
; FILE REFERENCE: 0575/62259/JPW/SHS
; CURRENT APPLICATION NUMBER: US/09/804,626
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-804-626-2

Query Match      11.6%; Score 50; DB 10; Length 518;
Best Local Similarity 100.0%; Pred. No. 2.2e-32;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAARKFERQHMDSPDLGTDDDDKAMADIGS 50
|||||
DB 116 MHHHHHSSGLVPRGSGMKETAARKFERQHMDSPDLGTDDDDKAMADIGS 165

RESULT 11
US-10-046-583A-3
; Sequence 3, Application US/10046583A
; Patent No. US20020168743A1
; GENERAL INFORMATION:
; APPLICANT: Rice, John
; APPLICANT: Kloti, Andreas
; APPLICANT: Crawford, John
; APPLICANT: Lanning, Beth
; APPLICANT: Stewart, Sandy
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: IDENTIFICATION OF MODULATORS OF DEOXYXYLULOSE 5-PHOSPHATE
; FILE OF INVENTION: SYNTHASE ACTIVITY
; FILE REFERENCE: 2037 US Divisional
; CURRENT APPLICATION NUMBER: US/10/046,583A
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/626,589
; PRIOR FILING DATE: 2000-07-27
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 824
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Residues 1-165 are from the thioredoxin sequence
; OTHER INFORMATION: found in the vector pET32 supplied by No. US20020168743Alagen.
; OTHER INFORMATION: Residues 166-824 represent the tDPS sequence from
; OTHER INFORMATION: Arabidopsis shown in SEQ ID NO:2.
US-10-046-583A-3

Query Match      11.6%; Score 50; DB 9; Length 824;
Best Local Similarity 100.0%; Pred. No. 3.2e-32;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSGMKETAARKFERQHMDSPDLGTDDDDKAMADIGS 50
|||||
DB 116 MHHHHHSSGLVPRGSGMKETAARKFERQHMDSPDLGTDDDDKAMADIGS 165

RESULT 12
US-09-978-756-1
; Sequence 1, Application US/09978756
; Patent No. US20020160017A1
; GENERAL INFORMATION:
; APPLICANT: Holder, Anthony
; APPLICANT: Birdsell, Berry
```

```
; APPLICANT: Feeney, James
; APPLICANT: Morgan, William
; APPLICANT: Syed, Shabih
; TITLE OF INVENTION: Malaria Vaccine
; FILE REFERENCE: 18396/1005
; CURRENT APPLICATION NUMBER: US/09/978,756
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: PCT/GB00/01558
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 09/311,817
; PRIOR FILING DATE: 1999-05-13
; PRIOR APPLICATION NUMBER: 2,271,451
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 9909072.2
; PRIOR FILING DATE: 1999-04-20
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-978-756-1

Query Match      10.7%; Score 46; DB 9; Length 96;
Best Local Similarity 100.0%; Pred. No. 8.8e-30;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

*QY 352 NSGCFRHLDERECKLLNFKQEGDKCVENPNPTCNENGGCDADA 397
|||||
DB 15 NSGCFRHLDERECKLLNFKQEGDKCVENPNPTCNENGGCDADA 60

RESULT 13
US-10-087-464-35
; Sequence 35, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chishti, Athar
; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuerong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses There
; FILE REFERENCE: S1237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 35
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-087-464-35

Query Match      10.7%; Score 46; DB 9; Length 114;
Best Local Similarity 100.0%; Pred. No. 1e-29;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 NSGCFRHLDERECKLLNFKQEGDKCVENPNPTCNENGGCDADA 397
|||||
DB 15 NSGCFRHLDERECKLLNFKQEGDKCVENPNPTCNENGGCDADA 60

RESULT 14
US-09-681-938-1
; Sequence 1, Application US/09681938
; Publication No. US20030003584A1
; GENERAL INFORMATION:
; APPLICANT: Adams, Darius
; TITLE OF INVENTION: Liposomal Vector Binding Protein for Hepatocyte DNA Delivery
; FILE REFERENCE: 29147
```

```

: APPLICANT: On, Steven
: APPLICANT: Liu, David
: APPLICANT: Goel, Vikas
: APPLICANT: Li, Xuerong
: TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses There
:
: FILE REFERENCE: S1237/7019
: CURRENT APPLICATION NUMBER: US/10/087,464
: CURRENT FILING DATE: 2002-03-01
: PRIOR APPLICATION NUMBER: US 06/272,930
: PRIOR FILING DATE: 2001-03-02
: NUMBER OF SEQ ID NOS: 59
:

```

```
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-087-464-10

Query Match      10.7%; Score 46; DB 9; Length 1639;
Best Local Similarity 100.0%; Pred. No. 9.4e-29;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 NSCCFRLHDERECKLLNYKQBGDKCVENPNTCTENNGGCCDADA 397
|||||
DB 1540 NSCCFRLHDERECKLLNYKQBGDKCVENPNTCTENNGGCCDADA 1585

RESULT 19
US-10-012-896-1011
; Sequence 1011, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davlin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepner, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Wantanabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012,896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1011
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-1011

Query Match      10.4%; Score 45; DB.9; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.7e-28;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHSSGLVPRGSMKETAARKFERQHMDSPDLGTDDDDKAM 45
|||||
DB 116 MHHHHHSSGLVPRGSMKETAARKFERQHMDSPDLGTDDDDKAM 160

RESULT 20
US-10-042-945-28
; Sequence 28, Application US/10042945
; Publication No. US20030045468A1
; GENERAL INFORMATION:
; APPLICANT: Filing, Steven P.
```

```
; APPLICANT: Foy, Teresa M.
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Wang, Aijun
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Sutherland, R. Alec
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY,
; TITLE OF INVENTION: DIAGNOSIS AND MONITORING OF BREAST CANCER
; FILE REFERENCE: 210121.479C3
; CURRENT APPLICATION NUMBER: US/10/042,945
; CURRENT FILING DATE: 2002-01-08
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-042-945-28

Query Match      7.4%; Score 32; DB 9; Length 55;
Best Local Similarity 100.0%; Pred. No. 1e-18;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GSGMKETAARKFERQHMDSPDLGTDDDDKAMA 46
|||||
DB 1 GSGMKETAARKFERQHMDSPDLGTDDDDKAMA 32

RESULT 21
US-09-757-417-28
; Sequence 28, Application US/09757417
; Patent No. US20020082216A1
; GENERAL INFORMATION:
; APPLICANT: Fanger, Gary R.
; APPLICANT: Foy, Theresa M.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY, DIAGNOSIS AND MONITORING OF BREAST CANCER
; FILE REFERENCE: 210121.479C1
; CURRENT APPLICATION NUMBER: US/09/757,417
; CURRENT FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-757-417-28

Query Match      7.4%; Score 32; DB 10; Length 55;
Best Local Similarity 100.0%; Pred. No. 1e-18;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GSGMKETAARKFERQHMDSPDLGTDDDDKAMA 46
|||||
DB 1 GSGMKETAARKFERQHMDSPDLGTDDDDKAMA 32

RESULT 22
US-09-872-712-1
; Sequence 1, Application US/09872712
; Publication No. US20030059461A1
; GENERAL INFORMATION:
; APPLICANT: Backer, Marina V.
; APPLICANT: Backer, Joseph M.
; TITLE OF INVENTION: MOLECULAR DELIVERY VEHICLE FOR DELIVERY
; TITLE OF INVENTION: OF SELECTED COMPOUNDS TO TARGETS
; FILE REFERENCE: 102131-200
; CURRENT APPLICATION NUMBER: US/09/872,712
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: 60/209,660
; PRIOR FILING DATE: 2000-06-05
```


REFERENCE/DOCKET NUMBER: 109.034US1
TELEPHONE: (612) 373-6900
TELEFAX: (612) 339-3061
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
US-09-096-749A-114

Query Match 3.5%; Score 15; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16
Db 5 HHHHHSSGLVPRGS 19

RESULT 27

US-10-060-275-5
Sequence 5, Application US/10060275
Publication No. US20030073828A1
GENERAL INFORMATION:
APPLICANT: DAI, ZIYU
APPLICANT: SHI, LIFANG
APPLICANT: HOOKER, BRIAN S.
TITLE OF INVENTION: EPIMERASE GENE AND USE THEREOF
FILE REFERENCE: 059440-0143
CURRENT APPLICATION NUMBER: US/10/060,275
CURRENT FILING DATE: 2002-06-11
PRIOR APPLICATION NUMBER: 60/265,311
PRIOR FILING DATE: 2001-02-01
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 21
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic amino
OTHER INFORMATION: acid sequence

US-10-060-275-5

Query Match 3.5%; Score 15; DB 9; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16
Db 5 HHHHHSSGLVPRGS 19

RESULT 28

US-09-736-084-98
Sequence 98, Application US/09736084
Patent No. US20020107211A1
GENERAL INFORMATION:

APPLICANT: THE ROCKEFELLER UNIVERSITY
TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING
NUCLEIC ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack

STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/09/736,084
APPLICATION NUMBER: 08/438,431
FILING DATE: May 10, 1995
APPLICATION NUMBER: 08/347,563
FILING DATE: NO. US20020107211A1
APPLICATION NUMBER: 08/292,345
FILING DATE: August 17, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-087 CIP21
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 98:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal
SEQUENCE DESCRIPTION: SEQ ID NO: 98:
US-09-736-084-98

Query Match 3.5%; Score 15; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16
Db 5 HHHHHSSGLVPRGS 19

RESULT 29

US-09-736-084-12
Sequence 12, Application US/09736084
Patent No. US20020107211A1
GENERAL INFORMATION:

APPLICANT: THE ROCKEFELLER UNIVERSITY
TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING
NUCLEIC ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/09/736,084
APPLICATION NUMBER: 08/438,431
FILING DATE: 13-Dec-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/438,431

; FILING DATE: May 10, 1995
; APPLICATION NUMBER: 08/347,563
; FILING DATE: NO. US20020107211a1ember 30, 1994
; APPLICATION NUMBER: 08/292,345
; FILING DATE: August 17, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-087 CIP21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-736-084-12

Query Match 3.5% Score 15; DB 10; Length 43;
Best Local Similarity 100.0%; Pred. No. 3.9e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16
Db 5 HHHHHSSGLVPRGS 19
|||||

RESULT 30
US-10-267-311-12
; Sequence 12, Application US/10267311
; Publication No. US20030050469A1
; GENERAL INFORMATION:
; APPLICANT: Siegel, Marvin
; APPLICANT: Chui, N. Randall
; APPLICANT: Mizzen, Lee A.
; TITLE OF INVENTION: INDUCTION OF A THI-LIKE RESPONSE IN VITRO
; FILE REFERENCE: 12071/002001
; CURRENT APPLICATION NUMBER: US/10/267,311
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US/09/613,303
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/143,757
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion sequence
US-10-267-311-12

Query Match 3.5% Score 15; DB 9; Length 121;
Best Local Similarity 100.0%; Pred. No. 9.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16
Db 5 HHHHHSSGLVPRGS 19
|||||

RESULT 31
US-09-981-286A-8
; Sequence 8, Application US/09981286A
; Publication No. US20020192799A1
; GENERAL INFORMATION:
; APPLICANT: Watowich, Stanley J.
; APPLICANT: Weaver, Scott C.

; APPLICANT: Davey, Robert A.
; TITLE OF INVENTION: Drug Discovery Methods
; FILE REFERENCE: 265.00260101
; CURRENT APPLICATION NUMBER: US/09/981,286A
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/240,187
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 124
; TYPE: PRT
; ORGANISM: BOS TAURUS
US-09-981-286A-8

Query Match 3.5% Score 15; DB 9; Length 124;
Best Local Similarity 100.0%; Pred. No. 9.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAKFERQHMD 33
Db 1 KETAAKFERQHMD 15
|||||

RESULT 32
US-09-876-348A-19
; Sequence 19, Application US/09876348A
; Patent No. US20020172951A1
; GENERAL INFORMATION:
; APPLICANT: Horwath, K. L. and Myers, K. L.
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
; FILE REFERENCE: RB-125-RI
; CURRENT APPLICATION NUMBER: US/09/876,348A
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/210,446
; PRIOR FILING DATE: 2000-06-08
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Word
; SEQ ID NO 19
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Tenebrio molitor
; OTHER INFORMATION: Mature Protein with His-tag, Clone 2.2
US-09-876-348A-19

Query Match 3.5% Score 15; DB 9; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16
Db 5 HHHHHSSGLVPRGS 19
|||||

RESULT 33
US-09-876-348A-23
; Sequence 23, Application US/09876348A
; Patent No. US20020172951A1
; GENERAL INFORMATION:
; APPLICANT: Horwath, K. L. and Myers, K. L.
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
; FILE REFERENCE: RB-125-RI
; CURRENT APPLICATION NUMBER: US/09/876,348A
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/210,446
; PRIOR FILING DATE: 2000-06-08
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Word
; SEQ ID NO 23
; LENGTH: 149
; TYPE: PRT

; ORGANISM: Tenebrio molitor
; OTHER INFORMATION: Mature Protein with His-tag, Clone 2.3
US-09-876-348A-23

Query Match 3.5%; Score 15; DB 9; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16
| | | | | | | | | | | | | | | | | | | | | |
Db 5 HHHHHSSGLVPRGS 19

RESULT 34

US-09-876-348A-27

; Sequence 27, Application US/09876348A

; Patent No. US20020172951A1

; GENERAL INFORMATION:

; APPLICANT: Horwath, K. L. and Myers, K. L.

; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
; Antifreeze Proteins and Method for Assaying Activity.

; FILE REFERENCE: RB-125-RI

; CURRENT APPLICATION NUMBER: US/09/876,348A

; CURRENT FILING DATE: 2001-08-09

; PRIOR APPLICATION NUMBER: 60/210,446

; PRIOR FILING DATE: 2000-06-08

; NUMBER OF SEQ ID NOS: 48

; SOFTWARE: Microsoft Word

; SEQ ID NO 27

; LENGTH: 149

; TYPE: PRT

; ORGANISM: Tenebrio molitor

; OTHER INFORMATION: Mature Protein with His-tag, Tm 13.17
US-09-876-348A-27

Query Match 3.5%; Score 15; DB 9; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16
| | | | | | | | | | | | | | | | | | | | | |
Db 5 HHHHHSSGLVPRGS 19

RESULT 35

US-09-876-348A-31

; Sequence 31, Application US/09876348A

; Patent No. US20020172951A1

; GENERAL INFORMATION:

; APPLICANT: Horwath, K. L. and Myers, K. L.

; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
; Antifreeze Proteins and Method for Assaying Activity.

; FILE REFERENCE: RB-125-RI

; CURRENT APPLICATION NUMBER: US/09/876,348A

; CURRENT FILING DATE: 2001-08-09

; PRIOR APPLICATION NUMBER: 60/210,446

; PRIOR FILING DATE: 2000-06-08

; NUMBER OF SEQ ID NOS: 48

; SOFTWARE: Microsoft Word

; SEQ ID NO 31

; LENGTH: 149

; TYPE: PRT

; ORGANISM: Tenebrio molitor

; OTHER INFORMATION: Mature Protein with His-tag, Clone 3.4
US-09-876-348A-31

Query Match 3.5%; Score 15; DB 9; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16
| | | | | | | | | | | | | | | | | | | | | |
Db 5 HHHHHSSGLVPRGS 19

RESULT 36

US-09-876-348A-35

; Sequence 35, Application US/09876348A

; Patent No. US20020172951A1

; GENERAL INFORMATION:

; APPLICANT: Horwath, K. L. and Myers, K. L.

; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
; Antifreeze Proteins and Method for Assaying Activity.

; FILE REFERENCE: RB-125-RI

; CURRENT APPLICATION NUMBER: US/09/876,348A

; CURRENT FILING DATE: 2001-08-09

; PRIOR APPLICATION NUMBER: 60/210,446

; PRIOR FILING DATE: 2000-06-08

; NUMBER OF SEQ ID NOS: 48

; SOFTWARE: Microsoft Word

; SEQ ID NO 35

; LENGTH: 149

; TYPE: PRT

; ORGANISM: Tenebrio molitor

; OTHER INFORMATION: Mature Protein with His-tag, Clone 3.9
US-09-876-348A-35

Query Match 3.5%; Score 15; DB 9; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16
| | | | | | | | | | | | | | | | | | | | | |
Db 5 HHHHHSSGLVPRGS 19

RESULT 37

US-09-876-348A-39

; Sequence 39, Application US/09876348A

; Patent No. US20020172951A1

; GENERAL INFORMATION:

; APPLICANT: Horwath, K. L. and Myers, K. L.

; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
; Antifreeze Proteins and Method for Assaying Activity.

; FILE REFERENCE: RB-125-RI

; CURRENT APPLICATION NUMBER: US/09/876,348A

; CURRENT FILING DATE: 2001-08-09

; PRIOR APPLICATION NUMBER: 60/210,446

; PRIOR FILING DATE: 2000-06-08

; NUMBER OF SEQ ID NOS: 48

; SOFTWARE: Microsoft Word

; SEQ ID NO 39

; LENGTH: 149

; TYPE: PRT

; ORGANISM: Tenebrio molitor

; OTHER INFORMATION: Mature protein with His-tag, Clone 7.5
US-09-876-348A-39

Query Match 3.5%; Score 15; DB 9; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHSSGLVPRGS 16
| | | | | | | | | | | | | | | | | | | | | |
Db 5 HHHHHSSGLVPRGS 19

RESULT 38

US-09-876-796A-19

; Sequence 19, Application US/09876796A

; Patent No. US20020173024A1

; GENERAL INFORMATION:

; APPLICANT: Horwath, K. L. and Easton, C. M.

; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
; Antifreeze Proteins and Method for Assaying Activity.

; FILE REFERENCE: RB-125-SEQ

; CURRENT APPLICATION NUMBER: US/09/876,796A
; CURRENT FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/210,446
; PRIOR FILING DATE: 2000-06-08
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Word
; SEQ ID NO 19
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Tenebrio molitor
; OTHER INFORMATION: Mature Protein with His-tag, Clone 2.2
US-09-876-796A-19

Query Match 3.5%; Score 15; DB 9; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16
| | | | | | | | | | | | | | | |
Db 5 HHHHHSSGLVPRGS 19

RESULT 39

US-09-876-796A-23
; Sequence 23, Application US/09876796A
; Patent No. US20020173024A1
; GENERAL INFORMATION:

; APPLICANT: Horwath, K. L. and Easton, C. M.
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
; TITLE OF INVENTION: Antifreeze Proteins and Method for Assaying Activity.

; FILE REFERENCE: RB-125-SEQ
; CURRENT APPLICATION NUMBER: US/09/876,796A

; CURRENT FILING DATE: 2001-06-07

; PRIOR APPLICATION NUMBER: 60/210,446

; PRIOR FILING DATE: 2000-06-08

; NUMBER OF SEQ ID NOS: 48

; SOFTWARE: Microsoft Word

; SEQ ID NO 23

; LENGTH: 149

; TYPE: PRT

; ORGANISM: Tenebrio molitor

; OTHER INFORMATION: Mature Protein with His-tag, Clone 2.3

US-09-876-796A-23

Query Match 3.5%; Score 15; DB 9; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16
| | | | | | | | | | | | | | | |
Db 5 HHHHHSSGLVPRGS 19

RESULT 40

US-09-876-796A-27
; Sequence 27, Application US/09876796A
; Patent No. US20020173024A1
; GENERAL INFORMATION:

; APPLICANT: Horwath, K. L. and Easton, C. M.
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
; TITLE OF INVENTION: Antifreeze Proteins and Method for Assaying Activity.

; FILE REFERENCE: RB-125-SEQ
; CURRENT APPLICATION NUMBER: US/09/876,796A

; CURRENT FILING DATE: 2001-06-07

; PRIOR APPLICATION NUMBER: 60/210,446

; PRIOR FILING DATE: 2000-06-08

; NUMBER OF SEQ ID NOS: 48

; SOFTWARE: Microsoft Word

; SEQ ID NO 27

; LENGTH: 149

; TYPE: PRT

; ORGANISM: Tenebrio molitor

; OTHER INFORMATION: Mature Protein with His-tag, Tm 13.17

US-09-876-796A-27

Query Match 3.5%; Score 15; DB 9; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HHHHHSSGLVPRGS 16
| | | | | | | | | | | | | | | |
Db 5 HHHHHSSGLVPRGS 19

Search completed: May 12, 2003, 10:35:42
Job time : 38 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 10:25:18 ; Search time 21 Seconds
(without alignments)
1973.045 Million cell updates/sec

Title: US-10-057-531A-2
Perfect score: 431
Sequence: 1 MHHHHHSSGLVPRGSMKE.....TCECTKPDSPFLDFGICSS 431

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	231	53.6	1726	1 SAZQGM	major merozoite su
2	190	44.1	1701	2 A54498	major merozoite su
3	154	35.7	1726	2 A45948	major merozoite su
4	145	33.6	651	2 S47282	merozoite surface
5	135	31.3	1701	2 A28868	major merozoite su
6	46	10.7	400	2 A45545	major merozoite su
7	46	10.7	1639	2 S05603	major merozoite su
8	24	5.6	1631	1 SAZQK1	major merozoite su
9	23	5.3	1640	2 A24594	probable major sur
10	15	3.5	124	1 NRBOB	pancreatic ribonuc
11	15	3.5	124	1 NRANE	pancreatic ribonuc
12	15	3.5	124	2 S08547	pancreatic ribonuc
13	15	3.5	124	2 JC5560	pancreatic ribonuc
14	15	3.5	150	1 NRBO	pancreatic ribonuc
15	14	3.2	30	2 A05004	pancreatic ribonuc
16	12	2.8	124	1 NRGN	pancreatic ribonuc
17	12	2.8	124	1 NRSH	pancreatic ribonuc
18	12	2.8	124	1 NRGT	pancreatic ribonuc
19	12	2.8	124	1 NRPRH	pancreatic ribonuc
20	12	2.8	124	2 S07141	pancreatic ribonuc
21	12	2.8	125	4 A47498	seminal ribonuclea
22	12	2.8	150	1 NRBOB	seminal ribonuclea
23	11	2.6	124	1 NRDER	pancreatic ribonuc
24	11	2.6	124	1 NRDEO	pancreatic ribonuc
25	11	2.6	124	1 NRDEF	pancreatic ribonuc
26	11	2.6	124	1 NRDEN	pancreatic ribonuc
27	11	2.6	124	1 NRKEN	pancreatic ribonuc
28	9	2.1	124	1 NRGF	pancreatic ribonuc
29	9	2.1	124	1 NRGM	pancreatic ribonuc

30	9	2.1	124	1 NRCMM	pancreatic ribonuc
31	9	2.1	124	1 NRCMB	pancreatic ribonuc
32	9	2.1	124	1 NRCZ	pancreatic ribonuc
33	9	2.1	124	1 NRHY	pancreatic ribonuc
34	9	2.1	128	1 NRHO	pancreatic ribonuc
35	9	2.1	128	1 NRCU	pancreatic ribonuc
36	9	2.1	128	1 NRPO	pancreatic ribonuc
37	9	2.1	130	2 S22808	pancreatic ribonuc
38	8	1.9	59	1 RSBLP5	integrase - Staphy
39	8	1.9	60	2 C64698	probable histidine
40	8	1.9	124	1 NRWB	pancreatic ribonuc
41	8	1.9	124	2 S08549	ribonuclease - dom
42	8	1.9	158	2 E86191	hypothetical prote
43	8	1.9	231	2 D64486	hypothetical prote
44	8	1.9	278	2 S75883	hypothetical prote
45	8	1.9	313	2 T28312	ORF MSV151 probabl
46	8	1.9	365	2 G81194	hypothetical prote
47	8	1.9	379	2 C86821	hypothetical prote
48	8	1.9	416	2 C71620	protein with Egl-1
49	8	1.9	582	2 A43412	semenogelin II pre
50	8	1.9	583	2 T04326	phosphoglucomutase
51	8	1.9	583	2 T04327	phosphoglucomutase
52	8	1.9	717	2 S38177	SSV7 protein homol
53	8	1.9	815	2 F86311	FlilA6.2 protein -
54	8	1.9	972	2 S35521	DNA topoisomerase
55	8	1.9	1356	2 S51389	ROM2 protein - yea
56	7	1.6	83	2 T16435	hypothetical prote
57	7	1.6	88	2 A69775	conserved hypothet
58	7	1.6	91	2 G86751	prophage p12 prote
59	7	1.6	92	2 F84620	similar to late em
60	7	1.6	109	2 C84525	hypothetical prote
61	7	1.6	112	2 E71167	hypothetical prote
62	7	1.6	120	2 F84577	hypothetical prote
63	7	1.6	122	2 H86481	hypothetical prote
64	7	1.6	140	2 E72505	hypothetical prote
65	7	1.6	162	1 D71206	hypothetical prote
66	7	1.6	168	2 B72635	hypothetical prote
67	7	1.6	173	2 A82935	type I restriction
68	7	1.6	182	2 B84387	hypothetical prote
69	7	1.6	190	2 A72575	hypothetical prote
70	7	1.6	195	2 A96998	CDP-diglyceride sy
71	7	1.6	209	2 G70091	ABC transporter (A
72	7	1.6	210	2 B70599	hypothetical prote
73	7	1.6	211	2 F69391	hypothetical prote
74	7	1.6	214	2 T47489	hypothetical prote
75	7	1.6	214	2 T08922	hypothetical prote
76	7	1.6	216	1 C64081	L-fucose-phospha
77	7	1.6	216	2 I53100	ehAND - mouse
78	7	1.6	219	2 T32443	hypothetical prote
79	7	1.6	226	2 S54843	glutaredoxin-like
80	7	1.6	227	2 T41464	probable Golgi mem
81	7	1.6	227	2 T46264	hypothetical prote
82	7	1.6	232	1 A25108	homeotic protein H
83	7	1.6	242	2 AF2874	hypothetical prote
84	7	1.6	242	2 G97650	ATPase AGR_C4396
85	7	1.6	245	2 A55456	transcription acti
86	7	1.6	249	2 F90582	methionine aminope
87	7	1.6	250	2 H97330	transcription regu
88	7	1.6	265	2 T49008	hypothetical prote
89	7	1.6	276	2 T22526	hypothetical prote
90	7	1.6	278	2 T35146	glutamate-binding
91	7	1.6	279	2 E69595	transcription acti
92	7	1.6	284	2 G97155	probable kinase [i
93	7	1.6	306	2 A88040	protein F47F6.1 [i
94	7	1.6	311	2 T33436	hypothetical prote
95	7	1.6	314	2 F07766	hypothetical prote
96	7	1.6	324	2 AB0403	probable sugar ABC
97	7	1.6	328	2 T19322	hypothetical prote
98	7	1.6	331	2 T37797	probable succinyl -
99	7	1.6	333	2 T52594	squamosa promoter
100	7	1.6	342	2 D72263	translation releas
101	7	1.6	343	2 T29547	hypothetical prote
102	7	1.6	348	2 T04618	heat shock protein

103	7	1.6	356	2	C91019	hypothetical prote	176	7	1.6	1019	2	T00117	dve protein - frui
104	7	1.6	358	2	E85863	hypothetical prote	177	7	1.6	1104	2	C72409	reverse gyrase - T
105	7	1.6	366	2	F81414	probable transmem	178	7	1.6	1122	2	T28130	hypothetical prote
106	7	1.6	373	2	B69860	conserved hypotet	179	7	1.6	1173	2	T30608	proteophosphoglyc
107	7	1.6	374	2	H88503	protein B0361.4 [l	180	7	1.6	1180	2	S69205	stripe a/b protein
108	7	1.6	382	2	C90407	conserved hypotet	181	7	1.6	1214	2	T47659	spliceosomal-like
109	7	1.6	385	2	S62824	hypothetical prote	182	7	1.6	1244	2	T19068	hypothetical prote
110	7	1.6	396	1	H70730	cytochrome P450 Rv	183	7	1.6	1306	2	S42659	acetyl-CoA carboxy
111	7	1.6	401	2	S53405	probable membrane	184	7	1.6	1308	2	T15280	hypothetical prote
112	7	1.6	405	2	E90154	primase (Amino-end	185	7	1.6	1339	2	A55301	1,3-beta-D-glucan-
113	7	1.6	418	2	G84546	probable tetracycl	186	7	1.6	1441	2	T13717	CRAG protein - fru
114	7	1.6	420	2	T39712	hypothetical prote	187	7	1.6	1446	2	T30916	carboxypeptidase D
115	7	1.6	427	2	T42516	hypothetical prote	188	7	1.6	1597	1	BVFFSL	sol protein, large
116	7	1.6	435	2	A31596	clathrin coat asse	189	7	1.6	1597	2	T08428	gene small optic l
117	7	1.6	435	2	T49327	clathrin-associate	190	7	1.6	1625	2	T02921	acetyl-CoA carboxy
118	7	1.6	435	2	G02088	assembly protein 5	191	7	1.6	1685	2	T02750	acetyl-CoA carboxy
119	7	1.6	438	2	G64962	shikimate transpor	192	7	1.6	1726	2	A39401	merozoite surface
120	7	1.6	438	2	B90976	shikimate transpor	193	7	1.6	1751	2	A45604	major blood-stage
121	7	1.6	438	2	H85822	probable transport	194	7	1.6	2311	2	T06161	acetyl-CoA carboxy
122	7	1.6	459	2	T45576	anthranilate N-hyd	195	7	1.6	2325	2	T02235	acetyl-CoA carboxy
123	7	1.6	466	2	E75201	pyridoxal phosphat	196	6	1.4	16	1	LFEC	his operon leader
124	7	1.6	473	2	T48985	hypothetical prote	197	6	1.4	16	1	C90981	his operon leader
125	7	1.6	476	2	S57963	methyl CpG binding	198	6	1.4	16	2	A85827	his operon leader
126	7	1.6	490	2	S52830	HMS1 protein - yea	199	6	1.4	21	2	T02227	Ig heavy chain CDR
127	7	1.6	500	1	E75342	IMP dehydrogenase	200	6	1.4	32	2	B82378	hypothetical prote
128	7	1.6	500	2	A55568	monocarboxylate tr	201	6	1.4	52	2	C90235	hypothetical prote
129	7	1.6	503	2	H87596	tryptophan halogen	202	6	1.4	54	2	A99983	H repeat-associat
130	7	1.6	513	2	D64495	hypothetical prote	203	6	1.4	55	2	H97965	degenerate transpo
131	7	1.6	513	2	S39691	UMP-hexose-1-phosp	204	6	1.4	56	2	A12134	hypothetical prote
132	7	1.6	513	2	B64431	nodulation factor	205	6	1.4	57	2	H64698	histidine and glut
133	7	1.6	515	1	T38946	phosphoprotein pho	206	6	1.4	57	2	F81308	probable coiled-co
134	7	1.6	528	2	B75364	extracellular solu	207	6	1.4	60	2	D84411	hypothetical prote
135	7	1.6	532	2	T15354	hypothetical prote	208	6	1.4	60	2	AC2576	hypothetical prote
136	7	1.6	548	2	A05032	rpcC protein homol	209	6	1.4	64	2	S57787	hypothetical prote
137	7	1.6	554	2	JE0303	propanediol dehydr	210	6	1.4	70	2	E96930	uncharacterized Fe
138	7	1.6	554	2	A56111	glycerol dehydrata	211	6	1.4	76	1	B64700	carbon storage reg
139	7	1.6	554	2	AC0760	hypothetical prote	212	6	1.4	76	2	E71820	probable carbon st
140	7	1.6	555	2	T32105	related to a-agglu	213	6	1.4	76	2	A97869	hypothetical prote
141	7	1.6	590	2	T49672	gene pointed prote	214	6	1.4	77	2	C70306	conserved hypotet
142	7	1.6	623	1	S33167	hypothetical prote	215	6	1.4	77	2	D71821	probable histidine
143	7	1.6	624	2	H90071	hypothetical prote	216	6	1.4	77	2	T16436	hypothetical prote
144	7	1.6	642	2	A81798	RNA polymerase sig	217	6	1.4	77	2	E97137	hypothetical prote
145	7	1.6	642	2	F81072	RNA polymerase sig	218	6	1.4	78	2	AB1162	flagellar switch p
146	7	1.6	659	2	T20753	hypothetical prote	219	6	1.4	78	2	AB1521	weakly flagellar s
147	7	1.6	666	2	T17396	vrp protein - Dic	220	6	1.4	80	2	C69204	hypothetical prote
148	7	1.6	682	2	A44493	serum-inducible ki	221	6	1.4	82	2	T01133	hypothetical prote
149	7	1.6	684	2	H96918	probable regulator	222	6	1.4	83	2	I46058	hypothetical prote
150	7	1.6	746	2	T28004	hypothetical prote	223	6	1.4	83	2	E72151	caldesmon - bovine
151	7	1.6	753	2	S48267	probable membrane	224	6	1.4	84	2	S33432	B13R protein - var
152	7	1.6	768	2	AB1085	hypothetical prote	225	6	1.4	84	2	C97194	hypothetical prote
153	7	1.6	792	2	B82752	penicillin binding	226	6	1.4	86	2	B26393	endonexin - bovine
154	7	1.6	818	1	JC4058	fibroblast growth	227	6	1.4	86	2	T27405	hypothetical prote
155	7	1.6	825	1	A60386	interleukin-4 rece	228	6	1.4	87	2	G83523	conserved hypotet
156	7	1.6	829	2	JC4583	fibroblast growth	229	6	1.4	88	2	B46264	thioredoxin 2 - sl
157	7	1.6	845	2	C82135	chitinase VC1952 [230	6	1.4	90	2	S78047	DNA-directed RNA p
158	7	1.6	869	2	A55384	transcription fact	231	6	1.4	91	2	F70142	glu-tRNA amidotran
159	7	1.6	876	2	D96558	probable protein k	232	6	1.4	92	2	T34146	hypothetical prote
160	7	1.6	885	2	G95980	probable nitrate r	233	6	1.4	93	2	F84585	hypothetical prote
161	7	1.6	896	2	T51891	hypothetical prote	234	6	1.4	93	2	T48043	hypothetical prote
162	7	1.6	897	2	C90561	hypothetical prote	235	6	1.4	93	2	H97840	hypothetical prote
163	7	1.6	913	2	D82885	multiple banded an	236	6	1.4	95	2	F71731	glutaredoxin 3 (gr
164	7	1.6	925	2	T16235	hypothetical prote	237	6	1.4	96	1	W4LE35	Ed protein - human
165	7	1.6	946	2	T16297	hypothetical prote	238	6	1.4	99	2	F90262	hypothetical prote
166	7	1.6	951	2	S52728	H+-exporting ATPas	239	6	1.4	101	2	B44971	hypothetical prote
167	7	1.6	956	2	A45066	H+-exporting ATPas	240	6	1.4	102	2	S09892	hypothetical prote
168	7	1.6	956	2	A43637	H+-exporting ATPas	241	6	1.4	102	2	T30119	hypothetical prote
169	7	1.6	958	2	S50751	H+-exporting ATPas	242	6	1.4	102	2	C97733	glutaredoxin 3 lim
170	7	1.6	956	2	T52412	H+-exporting ATPas	243	6	1.4	103	2	C84448	hypothetical prote
171	7	1.6	957	2	T03846	probable plasma me	244	6	1.4	105	2	B97700	thioredoxin [impor
172	7	1.6	960	2	T06688	H+-exporting ATPas	245	6	1.4	105	2	T49655	thioredoxin [impor
173	7	1.6	963	2	T12087	H+-exporting ATPas	246	6	1.4	106	2	S33357	thioredoxin - stre
174	7	1.6	989	2	T46183	zinc proteinase (E	247	6	1.4	107	2	S55137	hypothetical prote
175	7	1.6	1001	2	T28897	hypothetical prote	248	6	1.4	110	2	S55208	hypothetical prote

249	6	1.4	110	2	A71454	hypothetical prote	322	6	1.4	143	2	JC7223	lebecin-like prote
250	6	1.4	110	2	T17639	hypothetical prote	323	6	1.4	143	2	A9315	conserved hypotet
251	6	1.4	111	2	T16516	hypothetical prote	324	6	1.4	144	2	S44547	hypothetical prote
252	6	1.4	113	2	S08455	hypothetical prote	325	6	1.4	144	2	A84188	hypothetical prote
253	6	1.4	113	2	C81220	hypothetical prote	326	6	1.4	144	2	G64423	phospholipase A2 (
254	6	1.4	113	2	G81990	hypothetical prote	327	6	1.4	146	1	PSFGA	YU8H12.12 (import
255	6	1.4	114	2	S37150	asr2 protein - tom	328	6	1.4	146	2	C86187	hypothetical prote
256	6	1.4	114	2	T20095	hypothetical prote	329	6	1.4	147	2	E82923	hypothetical prote
257	6	1.4	115	2	H72583	hypothetical prote	330	6	1.4	149	2	G64414	hypothetical prote
258	6	1.4	116	2	G64433	hypothetical prote	331	6	1.4	150	2	S54517	hypothetical prote
259	6	1.4	116	2	G84032	hypothetical prote	332	6	1.4	150	2	E83057	hypothetical prote
260	6	1.4	117	2	G64335	hypothetical prote	333	6	1.4	150	2	S74649	hypothetical prote
261	6	1.4	119	2	A86605	iojap superfamily	334	6	1.4	150	2	C64319	hypothetical prote
262	6	1.4	119	2	T72020	iojap homolog - Ch	335	6	1.4	152	2	G72213	ferric uptake regu
263	6	1.4	119	2	A11591	B. subtilis PBX5 p	336	6	1.4	152	2	T51796	hypothetical prote
264	6	1.4	121	1	A36270	hemoglobin - Tetra	337	6	1.4	153	2	I53530	pancreatic ribonuc
265	6	1.4	122	2	F83167	hypothetical prote	338	6	1.4	153	2	A82687	conserved hypotet
266	6	1.4	122	2	T01558	auxin-induced prot	339	6	1.4	153	2	D97469	ynaf protein (AE00
267	6	1.4	123	2	B31596	pancreatic ribonuc	340	6	1.4	153	2	C97218	conserved membrane
268	6	1.4	124	1	NRWCB	pancreatic ribonuc	341	6	1.4	154	2	E82457	hypothetical prote
269	6	1.4	124	1	NRCPA	pancreatic ribonuc	342	6	1.4	154	2	E90588	conserved hypotet
270	6	1.4	124	1	NRCPA	pancreatic ribonuc	343	6	1.4	154	2	S40522	Huntington's disea
271	6	1.4	124	1	NRUI	pancreatic ribonuc	344	6	1.4	155	2	C86637	hypothetical prote
272	6	1.4	124	2	C70442	ribosomal protein	345	6	1.4	156	1	NRH01	pancreatic ribonuc
273	6	1.4	126	2	H96914	chemotaxis respons	346	6	1.4	156	2	B97836	bacterioferritin c
274	6	1.4	126	2	T03757	hypothetical prote	347	6	1.4	156	2	A98073	hypothetical prote
275	6	1.4	128	1	NROW2	pancreatic ribonuc	348	6	1.4	156	2	A87567	conserved hypotet
276	6	1.4	128	2	A33083	pancreatic ribonuc	349	6	1.4	157	2	AD2282	hypothetical prote
277	6	1.4	128	2	T15017	hypothetical prote	350	6	1.4	158	2	F69741	gltx 5'-region con
278	6	1.4	128	2	G81220	hypothetical prote	351	6	1.4	159	2	G90103	rubredoxin (import
279	6	1.4	130	2	D17707	thioredoxin (trxa)	352	6	1.4	161	2	E75043	conserved hypotet
280	6	1.4	130	2	T29887	hypothetical prote	353	6	1.4	161	2	F71117	conserved hypotet
281	6	1.4	130	2	S14983	extensin class I (354	6	1.4	161	2	S61624	probable membrane
282	6	1.4	131	2	A97791	nifu protein [limp	355	6	1.4	162	2	S05712	phycocyanin 3 alph
283	6	1.4	131	2	H95339	hypothetical prote	356	6	1.4	162	2	H83792	hypothetical prote
284	6	1.4	132	2	S66476	cytochrome P450 (C	357	6	1.4	162	2	F95165	hypothetical prote
285	6	1.4	132	2	I39193	gene HOXA1 protein	358	6	1.4	162	2	E98031	hypothetical prote
286	6	1.4	133	2	AB2421	hypothetical prote	359	6	1.4	163	1	F81938	hypothetical prote
287	6	1.4	133	2	B30242	stem cell protein	360	6	1.4	163	2	D71169	cytochrome c552 NM
288	6	1.4	135	2	S69464	hypothetical prote	361	6	1.4	164	1	TVBE11	transferring prote
289	6	1.4	135	2	H87008	conserved hypotet	362	6	1.4	164	2	S29550	rhizopuspepsin (EC
290	6	1.4	135	2	T49640	hypothetical prote	363	6	1.4	165	2	A81382	shikimate kinase (
291	6	1.4	136	1	GGICE3	globin CTF-III - m	364	6	1.4	165	2	S37679	hypothetical prote
292	6	1.4	136	1	GGICE3	hemoglobin III - m	365	6	1.4	166	2	S37738	hypothetical 18.6K
293	6	1.4	136	2	T22959	hypothetical prote	366	6	1.4	166	2	T31015	hypothetical prote
294	6	1.4	137	1	T44983	methylmalonyl-CoA	367	6	1.4	166	2	T39825	hypothetical prote
295	6	1.4	137	2	E72465	hypothetical prote	368	6	1.4	167	1	ORECCW	purine binding che
296	6	1.4	137	2	B69039	hypothetical prote	369	6	1.4	167	2	S20066	pancreatic-type ri
297	6	1.4	137	2	AF0087	flagellar switch p	370	6	1.4	167	2	A26143	chemotaxis protein
298	6	1.4	137	2	AE1277	E. coli Mutr prote	371	6	1.4	167	2	E90953	positive regulator
299	6	1.4	137	2	AE1640	cytochrome c552 NM	372	6	1.4	167	2	A85802	purine binding che
300	6	1.4	138	1	B81167	hypothetical prote	373	6	1.4	167	2	AD0746	hypothetical prote
301	6	1.4	138	2	AE1968	hypothetical prote	374	6	1.4	167	2	D70385	hypothetical prote
302	6	1.4	138	2	G69232	hypothetical prote	375	6	1.4	168	2	A95183	dihydrofolate redu
303	6	1.4	138	2	E90531	hypothetical prote	376	6	1.4	168	2	D98050	signal peptidase I
304	6	1.4	139	2	T33968	hypothetical prote	377	6	1.4	168	2	G70132	hypothetical prote
305	6	1.4	140	2	T06554	probable profilin	378	6	1.4	168	2	AG2257	probable GDP-L-fuc
306	6	1.4	140	2	A54523	histidine-rich pro	379	6	1.4	169	2	E90983	GDP-mannose mannos
307	6	1.4	140	2	B86683	prophage p11 prote	380	6	1.4	169	2	H85828	CT144 hypotet
308	6	1.4	140	2	C86800	prophage p13 prote	381	6	1.4	169	2	E86522	conserved hypotet
309	6	1.4	140	2	T15083	hypothetical prote	382	6	1.4	169	2	G81570	hypothetical prote
310	6	1.4	141	2	T15210	probable cytochrom	383	6	1.4	170	2	T31938	hypothetical prote
311	6	1.4	141	2	A11105	ribosomal protein	384	6	1.4	170	2	T03322	gene l15 protein -
312	6	1.4	141	2	A11467	ribosomal protein	385	6	1.4	171	2	A99415	hypothetical prote
313	6	1.4	141	2	T06553	probable profilin	386	6	1.4	171	2	G90532	hypothetical prote
314	6	1.4	141	2	B98145	hypothetical prote	387	6	1.4	172	2	T51065	hypothetical prote
315	6	1.4	142	2	AB3143	hypothetical prote	388	6	1.4	172	2	T34464	hypothetical prote
316	6	1.4	142	2	S28693	hypothetical prote	389	6	1.4	173	2	E64540	hypothetical prote
317	6	1.4	143	1	B64421	conserved hypotet	390	6	1.4	174	2	S01189	NADH2 dehydrogenas
318	6	1.4	143	2	T12144	hypothetical prote	391	6	1.4	174	2	B30020	NADH2 dehydrogenas
319	6	1.4	143	2	F86168	hypothetical prote	392	6	1.4	174	2	T02869	globulin.1 precurs
320	6	1.4	143	2	T43180	hypothetical prote	393	6	1.4	174	2	S38258	mannose-binding ie
321	6	1.4	143	2	T40426	hypothetical prote	394	6	1.4	174	2	F75606	hypothetical prote

395	6	1.4	175	2	JE0189	inorganic diphosph	468	6	1.4	206	2	H98265	hypothetical prote
396	6	1.4	175	2	F81437	formylmethionine d	469	6	1.4	207	2	H85070	hypothetical prote
397	6	1.4	175	2	C39141	transcription repr	470	6	1.4	208	2	F64244	ATP synthase B cha
398	6	1.4	175	2	S09784	hypothetical prote	471	6	1.4	208	2	S55930	het-c4 protein - p
399	6	1.4	175	2	S75258	hypothetical prote	472	6	1.4	208	2	D96543	unknown protein [i
400	6	1.4	176	2	T48255	hypothetical prote	473	6	1.4	208	2	T24446	hypothetical prote
401	6	1.4	176	2	G89977	conserved hypotet	474	6	1.4	210	1	STOMC	hypothetical prote
402	6	1.4	177	2	F81175	inorganic hypophos	475	6	1.4	210	2	T41553	somatotropin precu
403	6	1.4	177	2	G81930	probable inorganic	476	6	1.4	210	2	S28955	thymidylate kinase
404	6	1.4	178	2	D97160	uracil phosphoribo	477	6	1.4	210	2	S15186	dTMP kinase (EC 2.
405	6	1.4	178	2	T08444	hypothetical prote	478	6	1.4	210	2	S15118	somatotropin - chi
406	6	1.4	179	2	AD2427	hypothetical prote	479	6	1.4	210	2	JS0118	somatotropin precu
407	6	1.4	180	2	S08624	hydrogenase (EC 1.	480	6	1.4	210	2	S03709	somatotropin precu
408	6	1.4	180	2	H91075	formate hydrogenly	481	6	1.4	210	2	S06489	somatotropin II pr
409	6	1.4	180	2	G85920	formate hydrogenly	482	6	1.4	210	2	A31363	somatotropin precu
410	6	1.4	180	2	AD0846	formate hydrogenly	483	6	1.4	210	2	A23154	somatotropin I pre
411	6	1.4	180	2	H69115	probable orotate p	484	6	1.4	210	2	S15340	somatotropin precu
412	6	1.4	180	2	F84507	hypothetical prote	485	6	1.4	210	2	S08389	hypothetical prote
413	6	1.4	180	2	E84774	probable RING zinc	486	6	1.4	210	2	S27489	probable lipoprote
414	6	1.4	180	2	F83252	hypothetical prote	487	6	1.4	210	2	F90404	purine phosphoribo
415	6	1.4	181	2	G65024	Hydrogenase-4 comp	488	6	1.4	210	2	S77469	hypothetical prote
416	6	1.4	181	2	F91047	hydrogenase 4 Fe-S	489	6	1.4	211	2	B82901	hypothetical prote
417	6	1.4	181	2	B85892	hydrogenase 4 Fe-S	490	6	1.4	212	2	D72235	conserved hypotet
418	6	1.4	181	2	B84488	Athlia retroelemen	491	6	1.4	212	2	B82920	hypothetical prote
419	6	1.4	182	2	S19203	gene HOX2.8 protei	492	6	1.4	214	2	D90185	SSU ribosomal prot
420	6	1.4	182	2	AC1653	Listeria prophage	493	6	1.4	214	2	B70470	hypothetical prote
421	6	1.4	183	2	F71641	ATP synthase delta	494	6	1.4	215	1	WMBE21	UL14 protein - hum
422	6	1.4	183	2	T14653	hypothetical prote	495	6	1.4	215	2	S24717	pyroglutaryl-pepti
423	6	1.4	183	2	E69353	hypothetical prote	496	6	1.4	216	2	AB2582	glutamine amidotra
424	6	1.4	184	2	D87605	transcription regu	497	6	1.4	216	2	H69553	conserved hypotet
425	6	1.4	184	2	T2601	hypothetical prote	498	6	1.4	217	2	T37859	probable transcrip
426	6	1.4	185	2	JK0021	somatotropin - ski	499	6	1.4	217	2	T31940	hypothetical prote
427	6	1.4	185	2	T31939	hypothetical prote	500	6	1.4	217	2	B71203	hypothetical prote
428	6	1.4	185	2	D85040	hypothetical prote	501	6	1.4	217	2	S27829	gametocytogenesis
429	6	1.4	185	2	D71854	fkbp-type peptidyl	502	6	1.4	218	2	S58769	brain-specific pro
430	6	1.4	185	2	C64660	peptidyl-prolyl ci	503	6	1.4	218	2	C82828	hypothetical prote
431	6	1.4	187	2	G75273	conserved hypotet	504	6	1.4	219	2	S66947	hypothetical prote
432	6	1.4	188	2	H82933	hypothetical prote	505	6	1.4	220	2	B81333	probable membrane
433	6	1.4	189	2	A32840	hypothetical 22k p	506	6	1.4	221	2	H97363	amidotransferase h
434	6	1.4	189	2	T36736	probable secreted	507	6	1.4	221	2	B81253	probable peptide A
435	6	1.4	189	2	D71978	hypothetical prote	508	6	1.4	221	2	JC4761	recombination acti
436	6	1.4	189	2	E64527	hypothetical prote	509	6	1.4	222	2	H70192	endonuclease III (
437	6	1.4	189	2	AE1446	hypothetical prote	510	6	1.4	222	2	I39192	gene HOXA1 protein
438	6	1.4	190	2	B81386	probable scaffold	511	6	1.4	222	2	H82530	hypothetical prote
439	6	1.4	191	2	A64704	probable periplasm	512	6	1.4	222	2	E97538	phosphoglycolate p
440	6	1.4	191	2	A71814	hypothetical prote	513	6	1.4	222	2	AH2757	hydrolase [importe
441	6	1.4	191	2	E75132	molybdopterin-guan	514	6	1.4	224	2	S61386	lcmv protein - Leg
442	6	1.4	192	2	G71516	endopeptidase Clp	515	6	1.4	225	2	AG2427	hypothetical prote
443	6	1.4	192	2	D83273	conserved hypotet	516	6	1.4	225	2	G75448	conserved hypotet
444	6	1.4	192	2	T39367	hypothetical prote	517	6	1.4	225	2	F86498	pts IIA protein wi
445	6	1.4	193	2	D69212	conserved hypotet	518	6	1.4	225	2	A84786	probable cis-Golgi
446	6	1.4	195	2	T14965	hypothetical prote	519	6	1.4	225	2	G71533	probable pts IIA p
447	6	1.4	196	2	PD0004	self-incompatibili	520	6	1.4	225	2	F72124	pts IIA protein +
448	6	1.4	196	2	B81672	endopeptidase Clp	521	6	1.4	226	2	A32302	chemotaxis protein
449	6	1.4	196	2	S41271	ribosomal protein	522	6	1.4	227	2	B90400	hypothetical prote
450	6	1.4	196	2	S41263	ribosomal protein	523	6	1.4	228	2	AE2200	two-component resp
451	6	1.4	196	2	S41262	ribosomal protein	524	6	1.4	228	2	S51738	translin - human
452	6	1.4	196	2	S41257	ribosomal protein	525	6	1.4	228	2	F90130	hypothetical prote
453	6	1.4	196	2	I40130	outer surface prot	526	6	1.4	229	2	S42376	hypothetical prote
454	6	1.4	196	2	A11361	probable scaffold	527	6	1.4	230	2	B90525	protoporphirogen o
455	6	1.4	199	1	G64070	imidazoleglycerol-	528	6	1.4	230	2	C87200	conserved hypotet
456	6	1.4	199	2	AD2591	somatotropin precu	529	6	1.4	230	2	T47866	regulatory protein
457	6	1.4	199	2	H88108	protein C46E10.8 l	530	6	1.4	231	1	A32667	NAD(P)H2 dehydroge
458	6	1.4	200	1	A37146	ribosomal protein	531	6	1.4	231	1	TRPGTR	trypsin (EC 3.4.21
459	6	1.4	200	2	B35116	anthranilate synth	532	6	1.4	231	2	T06258	superoxide dismuta
460	6	1.4	200	2	C83519	anthranilate synth	533	6	1.4	231	2	T06801	probable superoxid
461	6	1.4	200	2	AC3509	dephospho-CoA kina	534	6	1.4	231	2	S41570	chalcone isomerase
462	6	1.4	201	2	I40550	hypothetical prote	535	6	1.4	231	2	D84934	50S ribosomal prot
463	6	1.4	202	2	E75053	hypothetical prote	536	6	1.4	231	2	A75103	phosphoglycolate p
464	6	1.4	204	2	C83060	probable ribosomal	537	6	1.4	232	2	B64442	hypothetical prote
465	6	1.4	205	2	E97135	ribonuclease D lim	538	6	1.4	232	2	T17553	hypothetical prote
466	6	1.4	206	2	T25384	conserved hypotet	539	6	1.4	233	2	T25295	protein gp33 [bact
467	6	1.4	206	2	AI3018		540	6	1.4	233	2	AB1366	

541	6	1.4	233	2	A69412	hypothetical prote	614	6	1.4	261	1	S00812	glucose 1-dehydrog
542	6	1.4	234	2	D84075	hypothetical prote	615	6	1.4	261	2	A33528	glucose 1-dehydrog
543	6	1.4	235	2	D89101	protein F25E5.8 [i	616	6	1.4	261	2	JS0385	glucose 1-dehydrog
544	6	1.4	236	2	T24882	hypothetical prote	617	6	1.4	261	2	I35853	glucose 1-dehydrog
545	6	1.4	237	2	S41512	Brn-3b protein - m	618	6	1.4	261	2	I40224	glucose 1-dehydrog
546	6	1.4	237	2	D83969	RNA polymerase spo	619	6	1.4	261	2	I40225	glucose 1-dehydrog
547	6	1.4	237	2	S73263	hypothetical prote	620	6	1.4	261	2	D79629	glucose 1-dehydrog
548	6	1.4	237	2	H75374	beta-phosphoglucom	621	6	1.4	261	2	A70666	probable modA prot
549	6	1.4	237	2	A71082	hypothetical prote	622	6	1.4	261	2	B42882	motility protein (
550	6	1.4	237	2	E90555	hypothetical prote	623	6	1.4	262	2	S02299	glucose 1-dehydrog
551	6	1.4	237	2	S28560	early light-induce	624	6	1.4	262	2	F84994	glutamate racemase
552	6	1.4	237	2	G75076	hypothetical prote	625	6	1.4	262	2	I64200	hypothetical prote
553	6	1.4	238	2	F89854	hypothetical prote	626	6	1.4	262	2	S44803	hypothetical prote
554	6	1.4	238	2	C37416	bloodstream-specif	627	6	1.4	263	2	S44803	F10E9.4 protein -
555	6	1.4	239	1	JU0083	transcription init	628	6	1.4	263	2	S01227	glucose 1-dehydrog
556	6	1.4	239	1	B39441	transcription init	629	6	1.4	263	2	G84547	hypothetical prote
557	6	1.4	239	2	T51270	hypothetical prote	630	6	1.4	264	2	H86147	TiN6.7 protein - A
558	6	1.4	239	2	E86346	F18F4.1 protein -	631	6	1.4	264	2	A31362	probable terminase
559	6	1.4	239	2	AE1317	hypothetical prote	632	6	1.4	265	2	D81315	hypothetical prote
560	6	1.4	240	2	S66275	proline-rich prote	633	6	1.4	265	2	T42419	hypothetical prote
561	6	1.4	240	2	G71968	hypothetical prote	634	6	1.4	265	2	T40259	hypothetical prote
562	6	1.4	240	2	B64540	conserved hypotet	635	6	1.4	266	2	C81254	hypothetical prote
563	6	1.4	240	2	AC1735	protein gp33 (Bact	636	6	1.4	266	2	E69746	conserved hypotet
564	6	1.4	240	2	F98065	hypothetical prote	637	6	1.4	267	2	B82541	ABC transporter pe
565	6	1.4	240	2	F87537	hypothetical prote	638	6	1.4	267	2	D84750	hypothetical prote
566	6	1.4	240	2	T45727	tonB protein, prob	639	6	1.4	268	2	A58446	Ig heavy chain v r
567	6	1.4	241	2	H84864	hypothetical prote	640	6	1.4	269	2	F96506	hypothetical prote
568	6	1.4	241	2	D64498	hypothetical prote	641	6	1.4	269	2	D82333	hypothetical prote
569	6	1.4	241	2	T22075	hypothetical prote	642	6	1.4	270	2	A26480	knob protein - mal
570	6	1.4	242	2	T31950	hypothetical prote	643	6	1.4	270	2	T41759	LEF-1 orf14 - Bomb
571	6	1.4	243	2	B29624	spherulin lb precu	644	6	1.4	271	2	H85774	hypothetical prote
572	6	1.4	244	2	S73022	probable membrane	645	6	1.4	272	2	T26640	hypothetical prote
573	6	1.4	244	2	T08438	hypothetical prote	646	6	1.4	272	2	H70335	hypothetical prote
574	6	1.4	244	2	A89752	protein C33E10.1 [647	6	1.4	272	2	T22562	hypothetical prote
575	6	1.4	245	2	H96932	hypothetical prote	648	6	1.4	274	1	A57691	NAD(P)H2 dehydroge
576	6	1.4	246	2	T38833	hypothetical prote	649	6	1.4	274	2	D91178	probable ARAC-type
577	6	1.4	246	2	C82541	hypothetical UPF00	650	6	1.4	274	2	E86024	probable ARAC-type
578	6	1.4	247	1	TRDG	ABC transporter Ar	651	6	1.4	274	2	S47736	probable transcript
579	6	1.4	247	2	C69761	trypsin (EC 3.4.21	652	6	1.4	274	2	A89816	hypothetical prote
580	6	1.4	247	2	G82202	ABC-type transport	653	6	1.4	274	2	H70900	hypothetical prote
581	6	1.4	247	2	E86250	hypothetical prote	654	6	1.4	275	2	T02334	hypothetical prote
582	6	1.4	248	1	B64462	protein F25C20.16	655	6	1.4	275	2	G91011	probable urease ac
583	6	1.4	248	2	T38025	dethiobiotin synth	656	6	1.4	275	2	A85856	probable elongatio
584	6	1.4	248	2	T33230	probable vacuolar	657	6	1.4	275	2	B64986	hypothetical 30.9
585	6	1.4	248	2	T22755	hypothetical prote	658	6	1.4	275	2	AH1924	hypothetical prote
586	6	1.4	248	2	B90345	hypothetical prote	659	6	1.4	276	2	T51437	hypothetical prote
587	6	1.4	248	2	F97414	hypothetical prote	660	6	1.4	276	2	B87203	probable antiloti
588	6	1.4	248	2	AB2632	hypothetical prote	661	6	1.4	276	2	A64516	hypothetical prote
589	6	1.4	248	2	T28412	hypothetical prote	662	6	1.4	277	2	S22494	rRNA N-glycosidase
590	6	1.4	249	2	T51687	myb-related transc	663	6	1.4	277	2	A37416	bloodstream-specif
591	6	1.4	249	2	C90526	conserved hypotet	664	6	1.4	277	2	T20065	hypothetical prote
592	6	1.4	251	2	E70521	probable phosphotr	665	6	1.4	278	2	T31942	hypothetical prote
593	6	1.4	253	2	A64456	hypothetical prote	666	6	1.4	278	2	T31935	hypothetical prote
594	6	1.4	253	2	T04642	hypothetical prote	667	6	1.4	279	2	T18829	hypothetical prote
595	6	1.4	253	2	S76719	hypothetical prote	668	6	1.4	279	2	S58402	synaptotagmin V -
596	6	1.4	253	2	T18978	hypothetical prote	669	6	1.4	279	2	D82281	ferric vibriobactl
597	6	1.4	253	2	F72710	hypothetical prote	670	6	1.4	279	2	G95117	Hemk protein (limp
598	6	1.4	254	2	AC1118	triophosphate is	671	6	1.4	280	2	T15715	hypothetical prote
599	6	1.4	254	2	AE1478	triophosphate is	672	6	1.4	280	2	B36144	hypothetical prote
600	6	1.4	254	2	S68481	hypothetical prote	673	6	1.4	280	2	S25326	uncharacterized pr
601	6	1.4	255	2	S41511	Brn-3a protein - m	674	6	1.4	280	2	D97100	uncharacterized pr
602	6	1.4	255	2	B71490	hypothetical prote	675	6	1.4	281	2	E70173	dimethyladenosine
603	6	1.4	256	2	A32017	beta-lactamase (EC	676	6	1.4	281	2	T45370	ribosomal protein
604	6	1.4	256	2	H75404	outer membrane pro	677	6	1.4	282	2	T31941	hypothetical prote
605	6	1.4	257	1	PNS02	beta-lactamase (EC	678	6	1.4	283	2	F86450	hypothetical prote
606	6	1.4	257	1	PNS2S	H+-transporing tw	679	6	1.4	284	2	F85306	probable regulator
607	6	1.4	257	1	PKQ6	transcription regu	680	6	1.4	284	2	S62839	sulfate transport
608	6	1.4	257	2	C96965	transcription anti	681	6	1.4	284	2	G75447	hypothetical prote
609	6	1.4	258	2	A82874	transcription anti	682	6	1.4	285	2	B64045	D-arabinitol 2-deh
610	6	1.4	259	2	T51679	myb-related transc	683	6	1.4	285	2	E84766	probable AT-hook D
611	6	1.4	259	2	A70359	hydrogenase expres	684	6	1.4	285	2	C84562	probable syntxin
612	6	1.4	259	2	C64481	conserved hypotet	685	6	1.4	286	2	H86664	outer membrane lip
613	6	1.4	259	2	E86300	protein F309.30 [i	686	6	1.4	286	2	T31494	hypothetical prote

687	1.4	287	2	T09035	hypothetical prote	760	1.4	308	2	AH1255	conserved hypothet
688	1.4	287	2	T39689	hypothetical prote	761	1.4	308	2	T19846	hypothetical prote
689	1.4	288	2	F72164	A9R protein - vari	762	1.4	308	2	G84210	ABC transporter, A
690	1.4	288	2	T28550	hypothetical prote	763	1.4	308	2	D70875	probable PE protei
691	1.4	288	2	I36848	A8R protein - vari	764	1.4	308	2	T24732	hypothetical prote
692	1.4	288	2	T37395	probable 33.6K pro	765	1.4	308	2	A97100	probable Fe-S oxid
693	1.4	288	2	A42518	A8R protein - vacc	766	1.4	308	2	H89839	hypothetical prote
694	1.4	288	2	E83946	pyruvate synthase	767	1.4	309	2	T29293	hypothetical prote
695	1.4	289	2	AH0060	bis(5'-nucleosyl)-	768	1.4	310	1	G64079	carbamate kinase (
696	1.4	289	2	F70845	hypothetical prote	769	1.4	310	2	A84142	L-lactate dehydrog
697	1.4	289	2	H83433	hypothetical prote	770	1.4	310	2	C83483	probable 2-hydroxy
698	1.4	289	2	E97044	probable integrase	771	1.4	310	2	T26710	hypothetical prote
699	1.4	290	1	B64657	conserved hypothet	772	1.4	311	2	A56235	transcription acti
700	1.4	290	2	C71859	hypothetical prote	773	1.4	311	2	T40921	hypothetical prote
701	1.4	290	2	D81359	probable integral	774	1.4	311	2	T18705	hypothetical prote
702	1.4	290	2	T21868	hypothetical prote	775	1.4	312	2	S71207	serine O-acetyltra
703	1.4	291	2	A71194	hypothetical prote	776	1.4	312	2	A87449	conserved hypothet
704	1.4	291	2	T51067	hypothetical prote	777	1.4	313	2	C83874	arsenical pump-dri
705	1.4	291	2	A11724	ABC transporter (A	778	1.4	313	2	T33391	hypothetical prote
706	1.4	292	2	F51171	transcription fact	779	1.4	313	2	S46690	hypothetical prote
707	1.4	292	2	AH1354	ABC transporter (A	780	1.4	313	2	T39212	hypothetical prote
708	1.4	293	2	AC2215	UTP-glucose-1-phos	781	1.4	313	2	H96037	alpha-galactoside
709	1.4	293	2	D90559	gtp-binding protei	782	1.4	314	2	F96527	protein F27115.20
710	1.4	293	2	A40644	transcription regu	783	1.4	314	2	T23301	hypothetical prote
711	1.4	294	2	A57478	serine O-acetyltra	784	1.4	315	2	T06806	proline rich prote
712	1.4	294	2	D90194	hypothetical prote	785	1.4	315	2	G70165	aldose reductase h
713	1.4	294	2	T24404	hypothetical prote	786	1.4	315	2	S75143	sensory transducti
714	1.4	294	2	E69759	hypothetical prote	787	1.4	315	2	JC7572	somite Maf1 protei
715	1.4	295	2	A12683	transcription regu	788	1.4	315	2	T17340	hypothetical prote
716	1.4	295	2	B32058	mcbb protein - Esc	789	1.4	316	1	NJBW41	toxin M1-1 precurs
717	1.4	295	2	B86371	hypothetical prote	790	1.4	316	2	G97465	lysr type transcri
718	1.4	295	2	B84747	hypothetical prote	791	1.4	316	2	T34353	G protein-coupled
719	1.4	295	2	T22039	hypothetical prote	792	1.4	316	2	T08694	hypothetical prote
720	1.4	295	2	AG2858	2-dehydro-3-deoxyg	793	1.4	316	2	I64205	transcription anti
721	1.4	296	1	S55795	arginase (EC 3.5.3	794	1.4	317	2	E90811	flagellar hook-fill
722	1.4	296	2	S21306	hypothetical prote	795	1.4	317	2	A85671	flagellar hook-fill
723	1.4	297	2	E70482	Na(+) dependent tr	796	1.4	317	2	S44022	flagellar hook-ass
724	1.4	297	2	S23737	proline-rich prote	797	1.4	317	2	G70222	conserved hypothet
725	1.4	298	2	D69351	hypothetical prote	798	1.4	318	2	A99668	hypothetical prote
726	1.4	298	2	S41469	homeotic protein M	799	1.4	318	2	D85518	hypothetical prote
727	1.4	298	2	G96773	hypothetical prote	800	1.4	318	2	C64754	yaqO protein - Esc
728	1.4	299	2	T33390	hypothetical prote	801	1.4	319	1	LUBO4	annexin IV - bovin
729	1.4	299	2	T15240	hypothetical prote	802	1.4	319	2	AG0742	high-affinity zinc
730	1.4	299	2	D81384	probable lipoprote	803	1.4	319	2	S57969	repB protein - Rhi
731	1.4	300	2	T42343	hypothetical prote	804	1.4	319	2	D97768	hypothetical prote
732	1.4	300	2	T33392	hypothetical prote	805	1.4	320	2	S09208	chorion protein s3
733	1.4	300	2	G69476	hypothetical prote	806	1.4	320	2	E90597	serine/threonine-p
734	1.4	300	2	T47460	hypothetical prote	807	1.4	320	2	G83839	hypothetical prote
735	1.4	301	2	AB0164	FecCD transport fa	808	1.4	321	2	D86375	probable aspartate
736	1.4	302	2	AE2520	hypothetical prote	809	1.4	321	2	T02987	myb-related protei
737	1.4	302	2	A55841	homeotic protein G	810	1.4	321	2	B82892	conserved hypothet
738	1.4	302	2	AC2018	hypothetical prote	811	1.4	322	2	F84945	glycine-tRNA ligas
739	1.4	303	2	A56837	homeotic protein M	812	1.4	322	2	A32329	complement C3 - Af
740	1.4	303	2	B49122	homeobox protein M	813	1.4	322	2	S13955	corticolliberin-bin
741	1.4	303	2	A48130	growth arrest-spec	814	1.4	322	2	T148207	corticolliberin-bin
742	1.4	304	2	B36716	prochlorophyllid	815	1.4	322	2	S13640	transcription fact
743	1.4	304	2	A55852	membrane-associate	816	1.4	323	2	T49529	hypothetical 36K p
744	1.4	304	2	B88746	protein C18F3.3 [i	817	1.4	323	2	S56558	hypothetical 36K p
745	1.4	304	2	S59414	hypothetical prote	818	1.4	323	2	I51751	homeotic protein o
746	1.4	304	2	AD2245	hypothetical prote	819	1.4	323	2	T48160	transcription fact
747	1.4	305	2	A56554	transcription fact	820	1.4	324	2	B85064	MYB-like protein [
748	1.4	305	2	I57039	genomic screen hom	821	1.4	324	2	D86834	transcription regu
749	1.4	305	2	T24129	hypothetical prote	822	1.4	324	2	A70111	hypothetical prote
750	1.4	305	2	T25117	hypothetical prote	823	1.4	324	2	T28032	hypothetical prote
751	1.4	306	2	B64073	ribokinase (EC 2.7	824	1.4	325	2	T39169	probable oxidoredu
752	1.4	306	2	B25394	SURF-1 protein - m	825	1.4	325	2	E90168	DNA modification m
753	1.4	306	2	T09067	hypothetical prote	826	1.4	325	2	B90532	fatty acid/phospho
754	1.4	307	2	S38152	hypothetical prote	827	1.4	325	2	A37159	translation elonga
755	1.4	307	2	A45581	distal-less homeob	828	1.4	325	2	E86533	probable leader 19
756	1.4	307	2	T34143	hypothetical prote	829	1.4	325	2	E81578	conserved hypothet
757	1.4	308	2	E70392	cation efflux syst	830	1.4	325	2	A81713	conserved hypothet
758	1.4	308	2	G90057	conserved hypothet	831	1.4	326	1	S44259	sucrose operon reg
759	1.4	308	2	AE1618	conserved hypothet	832	1.4	326	2	D83483	probable metal tra

833	6	1.4	326	2	T19070	hypothetical prote	906	1.4	351	1	KGZQHL	histidine-rich gly
834	6	1.4	328	2	S16573	plastoquinol-plast	907	1.4	351	2	T32717	hypothetical prote
835	6	1.4	328	2	D72393	oligopeptide ABC t	908	1.4	351	2	T20270	hypothetical prote
836	6	1.4	328	2	S45998	hypothetical prote	909	1.4	351	2	T23423	hypothetical prote
837	6	1.4	328	2	T25231	hypothetical prote	910	1.4	352	2	T47820	hypothetical prote
838	6	1.4	328	2	G02469	homeotic protein b	911	1.4	353	2	JC7128	pleckstrin 2 - mou
839	6	1.4	329	1	S55321	cysteine synthase	912	1.4	353	2	H88939	protein C05E4.13 [
840	6	1.4	329	2	F91290	hypothetical prote	913	1.4	354	1	RGXLOA	GTP-binding regula
841	6	1.4	329	2	A86132	hypothetical prote	914	1.4	354	1	RGFFO2	GTP-binding regula
842	6	1.4	329	2	T45972	hypothetical prote	915	1.4	354	2	A61035	hypothetical prote
843	6	1.4	330	1	Q0BEP7	gene 62 protein -	916	1.4	354	2	A90012	homeotic protein o
844	6	1.4	330	2	C90090	ATP(GTP)-binding p	917	1.4	354	2	S39406	hypothetical prote
845	6	1.4	331	2	A30242	homeotic protein E	918	1.4	354	2	G96795	G protein alpha ch
846	6	1.4	331	2	S78452	POU-domain protein	919	1.4	355	2	T50479	conserved hypotet
847	6	1.4	332	1	DEHULM	L-lactate dehydrog	920	1.4	355	2	A70185	probable glucose-1
848	6	1.4	332	1	B39136	flagellar motor sw	921	1.4	355	2	A72589	otxl protein - mou
849	6	1.4	332	2	T20107	hypothetical prote	922	1.4	355	2	S35345	homeodomain protel
850	6	1.4	332	2	H72313	transcription regu	923	1.4	355	2	I58547	hypothetical prote
851	6	1.4	332	2	JH0465	homeotic protein T	924	1.4	355	2	C70194	hypothetical prote
852	6	1.4	332	2	A84523	probable steroid s	925	1.4	355	2	T15759	probable fibmbrial
853	6	1.4	334	2	A97020	isocitrate dehydro	926	1.4	355	2	AF0955	hypothetical prote
854	6	1.4	334	2	T22427	hypothetical prote	927	1.4	356	2	H72214	hypothetical prote
855	6	1.4	334	2	F97273	catabolite control	928	1.4	356	2	T48354	hypothetical prote
856	6	1.4	334	2	B96810	unknown protein Tl	929	1.4	357	2	S63146	hypothetical prote
857	6	1.4	334	2	T17213	hypothetical prote	930	1.4	358	2	T20286	hypothetical prote
858	6	1.4	334	2	E97635	hypothetical prote	931	1.4	358	2	S57176	hypothetical prote
859	6	1.4	335	2	C75119	deoxyhypusine synt	932	1.4	358	2	G69754	hypothetical prote
860	6	1.4	335	2	G01448	homeobox protein H	933	1.4	359	2	D81598	hypothetical prote
861	6	1.4	336	2	AC1979	ferric iron-bindin	934	1.4	359	2	C72068	hypothetical prote
862	6	1.4	336	2	B86463	AIgI-like protein,	935	1.4	359	2	H86555	hypothetical prote
863	6	1.4	338	2	S38030	suppressor protein	936	1.4	359	2	S14283	transcription fact
864	6	1.4	338	2	D98233	hypothetical prote	937	1.4	359	2	T21705	hypothetical prote
865	6	1.4	338	2	AG3052	DNA topoisomerase	938	1.4	360	2	C70773	threonine synthase
866	6	1.4	339	1	LUBO36	annexin II - bovin	939	1.4	360	2	D71323	conserved hypotet
867	6	1.4	339	1	LUMS36	annexin II - mouse	940	1.4	360	2	AE1315	histidinol-phospha
868	6	1.4	339	2	LUCH2	transcription regu	941	1.4	361	2	AE1687	hypothetical prote
869	6	1.4	339	2	AH2127	probable membrane	942	1.4	361	2	C90603	conserved hypotet
870	6	1.4	339	2	S57152	protein F10B6.6 [i	943	1.4	361	2	A82887	probable peptidyl-
871	6	1.4	339	2	B86281	uncharacterized pr	944	1.4	362	2	T38464	transcription term
872	6	1.4	339	2	E97084	hypothetical prote	945	1.4	362	2	G81390	probable endo-1,3-
873	6	1.4	339	2	D91016	probable protein w	946	1.4	363	2	G96807	hypothetical prote
874	6	1.4	340	2	T51386	VSG expression sit	947	1.4	364	2	T22807	hypothetical prote
875	6	1.4	340	2	A34433	D-sterospecific a	948	1.4	364	2	T30084	hypothetical prote
876	6	1.4	340	2	AD3602	glyceroldehyde-3-p	949	1.4	364	2	T02801	probable membrane
877	6	1.4	341	1	DEKWG3	protein gp4-2 [imp	950	1.4	365	2	T11905	NADH2 dehydrogenas
878	6	1.4	341	1	DEKWG3	metalloproteinase	951	1.4	365	2	JC4027	glutamate-ammonia
879	6	1.4	341	2	A89491	probable metallopr	952	1.4	365	2	AD1634	X-Pro dipeptidase
880	6	1.4	341	2	T51957	hypothetical prote	953	1.4	365	2	S43780	peridin-chloroph
881	6	1.4	342	2	G84885	hypothetical prote	954	1.4	365	2	H87288	conserved hypotet
882	6	1.4	342	2	T18474	hypothetical prote	955	1.4	366	2	T34700	probable transcrip
883	6	1.4	342	2	AD2032	hypothetical prote	956	1.4	366	2	G70230	conserved hypotet
884	6	1.4	343	2	D89780	hypothetical prote	957	1.4	366	2	A99842	dihydroxyacetone k
885	6	1.4	343	2	S52009	farnesyl-pyrophosp	958	1.4	366	2	G85699	glucokinase (EC 2.
886	6	1.4	343	2	A84964	biotin synthase (E	959	1.4	366	2	E64866	hypothetical prote
887	6	1.4	344	1	A40005	hoscycamine (6S)-d	960	1.4	367	2	D84548	hypothetical prote
888	6	1.4	344	2	H71825	probable histidine	961	1.4	367	2	AH0688	probable bacteriop
889	6	1.4	344	2	T00695	hypothetical prote	962	1.4	368	2	H96712	probable DNA-bindi
890	6	1.4	345	1	S75490	probable hydro-ly	963	1.4	369	1	TVFVAF	transforming prote
891	6	1.4	345	2	F81751	conserved hypotet	964	1.4	369	2	B84356	AAA-type ATPase [1
892	6	1.4	345	2	S55377	urPAB protein prec	965	1.4	369	2	T40905	WD repeat protein
893	6	1.4	346	2	A03308	dnak-type molecula	966	1.4	369	2	D81261	probable ABC trans
894	6	1.4	346	2	S66958	UFEL protein - yea	967	1.4	370	2	I57555	c-Maf protein - mo
895	6	1.4	347	2	T09140	serine O-acetyltra	968	1.4	370	2	T31743	hypothetical prote
896	6	1.4	347	2	H91044	hypothetical prote	969	1.4	370	2	E97316	hypothetical prote
897	6	1.4	347	2	C85889	hypothetical prote	970	1.4	372	2	T09617	response regulator
898	6	1.4	347	2	G97256	spore photoproduct	971	1.4	373	2	T10967	recf3 protein - Str
899	6	1.4	349	2	E95858	conserved hypotet	972	1.4	373	2	T52182	probable transcrip
900	6	1.4	349	2	T19134	hypothetical prote	973	1.4	373	2	T20723	hypothetical prote
901	6	1.4	349	2	A96839	F23A5.2(form2) [im	974	1.4	374	1	A69954	conserved hypotet
902	6	1.4	349	2	B86769	GTP cyclohydrolase	975	1.4	374	1	G69359	conserved hypotet
903	6	1.4	349	2	AH2382	hypothetical prote	976	1.4	374	2	G64319	hypothetical prote
904	6	1.4	350	2	B82713	lipase modulator x	977	1.4	374	2	AB1943	hypothetical prote
905	6	1.4	350	2	A30046	homeotic protein r	978	1.4	374	2	H64500	hypothetical prote

979 6 1.4 374 2 AG3337 transcription regu
980 6 1.4 375 1 DECHAI alcohol dehydrogen
981 6 1.4 375 1 A35837 alcohol dehydrogen
982 6 1.4 375 2 T08572 hypothetical prote
983 6 1.4 375 2 C71441 probable SCARECROW
984 6 1.4 375 2 B82964 glycosyltransferas
985 6 1.4 375 2 A46390 CAMP receptor.subt
986 6 1.4 375 2 S49767 heat shock protein
987 6 1.4 376 1 C64135 ribonucleoside-dip
988 6 1.4 376 1 RGBY43 cell division cont
989 6 1.4 376 2 B72321 conserved hypothet
990 6 1.4 376 2 A49077 transcription init
991 6 1.4 377 2 T16763 hypothetical prote
992 6 1.4 377 2 AH3328 NAD(FAD)-utilizing
993 6 1.4 377 2 F82105 hypothetical prote
994 6 1.4 378 2 A49337 alanine dehydrogen
995 6 1.4 378 2 B90704 H repeat-associate
996 6 1.4 378 2 E85554 probable receptor
997 6 1.4 378 2 H83019 probable glycosyl
998 6 1.4 378 2 T21651 hypothetical prote
999 6 1.4 379 2 H89876 phosphoribosylamin
1000 6 1.4 379 2 S55245 polyubiquitin 5 -

ALIGNMENTS

RESULT 1
SAZQCM
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (s
N:Alternate names: 195K glycoprotein
C:Species: Plasmodium falciparum
C>Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000
C:Accession: A23386; S06361
R:Weber, J.L.; Leininger, W.M.; Lyon, J.A.
Nucleic Acids Res. 14, 3311-3323, 1986
A:Title: Variation in the gene encoding a major merozoite surface antigen of the human m
A:Reference number: A23386; MUID:86205236; PMID:3517809
A:Accession: A23386
A:Molecule type: DNA
A:Residues: 1-1104 <WEB1>
R:Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Wolff, R.
Nucleic Acids Res. 16, 1206, 1988
A:Title: Merozoite surface protein sequence from the Camp strain of the human malaria pa
A:Reference number: S06361; MUID:88143999; PMID:3278296
A:Accession: S06361
A:Molecule type: DNA
A:Residues: 1104-1726 <WEB2>
A:CROSS-references: EMBL:X03831
C:Comment: The merozoite stages of different strains have strain-specific surface antige
C:Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The me
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1726/Product: major merozoite surface antigen #status predicted <MAT>
F:67-87,91-96,100-105,109-120/Region: 3-residue repeats (S-G-T)
F:757-765/Region: 3-residue repeats (T-E-E)
F:133,272,501,567,638,827,839,924,944,990,1016,1114,1221,1613,1658/Binding site: carbohy

Query Match 53.6%; Score 231; DB 1; Length 1726;
Best Local Similarity 100.0%; Pred. No. 2.9e-216;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 58 AISVTMDNLGFEYDVIYKPLAGVYRSLLKQIEKNFTFNLDNLNLDILNSRLKRRKY 117
|||||
Db 1333 AISVTMDNLGFEYDVIYKPLAGVYRSLLKQIEKNFTFNLDNLNLDILNSRLKRRKY 1392
|||||
Qy 118 FLDVLESLDMQFKHSSNEYIIDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 177
|||||
Db 1393 FLDVLESLDMQFKHSSNEYIIDSFKLLNSEQKNTLLSKYIKESVENDIKFAQEGIS 1452
|||||
Qy 178 YEKVLAKYKDDLESIKKVIKEEKFPPSPPTTPSPAKTDEQKESKFLPFLTNIETL 237

Db 1453 YEKVLAKYKDDLESIKKVIKEEKFPPSPPTTPSPAKTDEQKESKFLPFLTNIETL 1512
|||||
Qy 238 YNNLVNKIDDDYLLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFN 288*
|||||
Db 1513 YNNLVNKIDDDYLLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFN 1563
|||||
RESULT 2
A54498
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
C:Accession: A54498
R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V
Mol. Biochem. Parasitol. 27, 291-302, 1988
A:Title: Variation in the precursor to the major merozoite surface antigens of Plasm
A:Reference number: A54498; MUID:88142999; PMID:2449612
A:Accession: A54498
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1701 <PET>
A:CROSS-references: GB:M19143; NID:g160412; PIDN:AAA29653.1; PID:g160413
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 44.1%; Score 190; DB 2; Length 1701;
Best Local Similarity 100.0%; Pred. No. 2.5e-176;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 99 TFNLNLDILNSRLKRRKYFDLVLESLDMQFKHSSNEYIIDSFKLLNSEQKNTLLKSY 158
|||||
Db 1349 TFNLNLDILNSRLKRRKYFDLVLESLDMQFKHSSNEYIIDSFKLLNSEQKNTLLKSY 1408
|||||
Qy 159 KYIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEEKFPPSPPTTPSPAKT 218
|||||
Db 1409 KYIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEEKFPPSPPTTPSPAKT 1468
|||||
Qy 219 DEQKESKFLPFLTNIETLYNNLVNKIDDDYLLINLAKINDCNVEKDEAHVKITKLSDLKA 278
|||||
Db 1469 DEQKESKFLPFLTNIETLYNNLVNKIDDDYLLINLAKINDCNVEKDEAHVKITKLSDLKA 1528
|||||
Qy 279 IDDKIDLKFN 288
|||||
Db 1529 IDDKIDLKFN 1538
|||||

RESULT 3
A45948
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jun-2000
C:Accession: A45948
R:Chang, S.P.; Kramer, K.J.; Yamaga, K.M.; Kato, A.; Case, S.E.; Siddiqui, W.A.
Exp. Parasitol. 67, 1-11, 1988
A:Title: Plasmodium falciparum: gene structure and hydropathy profile of the major me
A:Reference number: A45948; MUID:89005525; PMID:3049134
A:Accession: A45948
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1726 <CHA>
A:CROSS-references: GB:M37213
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 35.7%; Score 154; DB 2; Length 1726;
Best Local Similarity 100.0%; Pred. No. 3.1e-141;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 58 AISVTMDNLGFEYDVIYKPLAGVYRSLLKQIEKNFTFNLDNLNLDILNSRLKRRKY 117
|||||
Db 1333 AISVTMDNLGFEYDVIYKPLAGVYRSLLKQIEKNFTFNLDNLNLDILNSRLKRRKY 1392
|||||

QY 118 FLDVLESDLMQPKHSSNYYIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 177
|||||
Db 1393 FLDVLESDLMQPKHSSNYYIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 1452
QY 178 YVEKVLAKYKDDLESIKKVIKEEKEKFPSPPTT 211
|||||
Db 1453 YVEKVLAKYKDDLESIKKVIKEEKEKFPSPPTT 1486

RESULT 4
S47282
merozoite surface antigen 1 - malaria parasite (Plasmodium falciparum) (strain RO-71) (f
C:Species: Plasmodium falciparum
A:Variety: strain RO-71
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 09-Jun-2000
R:Tolle, R.; Bujard, H.; Cooper, J.A.
submitted to the EMBL Data Library, July 1994
A:Description: Plasmodium falciparum: recombination within the C-terminal region of mero
A:Reference number: S47282
A:Accession: S47282
A:Molecule type: DNA
A:Residues: 1-651 <TO>
A:Cross-references: EMBL:Z35329; NID:g535257; PIDN:CAA84558.1; PID:g535258
A:Experimental source: strain RO-71
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen

Query Match 33.6%; Score 145; DB 2; Length 651;
Best Local Similarity 100.0%; Pred. No. 7.9e-133; Mismatches 0; Indels 0; Gaps 0;
Matches 145; Conservative 0;

QY 58 AISVTMDNLTSGFENEYDVIYKPLAGVYRSLKQTEKNFTFNLDNLNSRLKRRKY 117
|||||

Db 259 AISVTMDNLTSGFENEYDVIYKPLAGVYRSLKQTEKNFTFNLDNLNSRLKRRKY 318
|||||

QY 118 FLDVLESDLMQPKHSSNYYIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 177
|||||

Db 319 FLDVLESDLMQPKHSSNYYIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGIS 378
|||||

QY 178 YVEKVLAKYKDDLESIKKVIKEEKE 202
|||||

Db 379 YVEKVLAKYKDDLESIKKVIKEEKE 403
|||||

RESULT 5
A26868
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
C:Species: Plasmodium falciparum
C:Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000
R:Tanabe, K.; Mackay, M.; Scaife, J.G.
J. Mol. Biol. 195, 273-287, 1987
A:Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium
A:Reference number: A26868; MUID:88011243; PMID:3079521
A:Accession: A26868
A:Molecule type: DNA
A:Residues: 1-1701 <TAN>
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1701/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 31.3%; Score 135; DB 2; Length 1701;
Best Local Similarity 100.0%; Pred. No. 1e-122;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 LKSKYKYIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEEKEKFPSPPTT 213
|||||

Db 1404 LKSKYKYIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEEKEKFPSPPTT 1463
|||||

QY 214 SPAKTEQKESKFLPFLTNIETLYNNLVNKIDYLLINLAKINCENVEKDEAHVKITKL 273
|||||

Db 1464 SPAKTEQKESKFLPFLTNIETLYNNLVNKIDYLLINLAKINCENVEKDEAHVKITKL 1523
QY 274 SDLKATDDKIDLFKN 288
|||||
Db 1524 SDLKATDDKIDLFKN 1538
|||||

RESULT 6
A45545
major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (fragments
C:Species: Plasmodium falciparum
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
A:Accession: A45545
R:Blackman, M.J.; Ling, I.T.; Nicholls, S.C.; Holder, A.A.
Mol. Biochem. Parasitol. 49, 29-33, 1991
A:Title: Proteolytic processing of the Plasmodium falciparum merozoite surface protei
A:Reference number: A45545; MUID:92131048; PMID:1775158
A:Accession: A45545
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-400 <BLA>
A:Note: Sequence extracted from NCBI backbone (NCBIN:77612, NCBIP:77621)
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

Query Match 10.7%; Score 46; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 1.5e-36; Mismatches 0; Indels 0; Gaps 0;
Matches 46; Conservative 0;

QY 352 NSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNNGCGDADA 397
|||||

Db 301 NSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNNGCGDADA 346
|||||

RESULT 7
S05603
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N:Alternate names: gp195 surface antigen
C:Species: Plasmodium falciparum
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
A:Accession: S05603; S04850
R:Myler, P.J.
submitted to the EMBL Data Library, April 1989.
A:Reference number: S05603
A:Accession: S05603
A:Molecule type: mRNA
A:Residues: 1-1639 <MYL>
A:Cross-references: EMBL:X15063; NID:g9896; PIDN:CAA33163.1; PID:g9897
Nucleic Acids Res. 17, 5401, 1989
A:Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from pl
A:Reference number: S04850; MUID:89345116; PMID:2668887
A:Accession: S04850
A:Molecule type: mRNA
A:Residues: 1504-1639 <MYL2>
A:Cross-references: EMBL:X15063
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1639/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 10.7%; Score 46; DB 2; Length 1639;
Best Local Similarity 100.0%; Pred. No. 5.2e-36; Mismatches 0; Indels 0; Gaps 0;
Matches 46; Conservative 0;

QY 352 NSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNNGCGDADA 397
|||||

Db 1540 NSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCENNNGCGDADA 1585
|||||

RESULT 8
SAZQK1
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
|||||

C:Species: Plasmodium falciparum
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 09-Jun-2000
C:Accession: A25120
R:Mackay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scalfe, J.; Certa, U.; Stunnenberg, H.; BEMBO J. 4, 3823-3829, 1985
A:Title: Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum
A:Reference number: A91030; MUID:86136024; PMID:3004972
A:Accession: A25120
A:Molecule type: DNA
A:Residues: 1-1631 <MAC>
C:Comment: The merozoite stages of different strains have strain-specific surface antigens
C:Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The merozoite stage is the most abundant in the blood.
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1631/Product: major merozoite surface antigen #status predicted <MAT>
F:67-84/Region: 3-residue repeats (S-G-T/P)
F:1614-1631/Domain: membrane anchor #status predicted <MBN>
F:97,259,755,759,835,911,955,1049,1156,1163,1436,1563/Binding site: carbohydrate (Asn) (Glycosylation site)
Query Match 5.6%; Score 24; DB 1; Length 1631;
Best Local Similarity 100.0%; Pred. No. 1.4e-14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 352 NSGCFRHLDERECKCLLNKYQEG 375
DB 1532 NSGCFRHLDERECKCLLNKYQEG 1555
RESULT 9
A24594
probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jun-2000
C:Accession: A24594
R:Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholls, Nature 317, 270-273, 1985
A:Title: Primary structure of the precursor to the three major surface antigens of Plasmodium falciparum
A:Reference number: A24594; MUID:86014355; PMID:2995820
A:Accession: A24594
A:Molecule type: DNA
A:Residues: 1-1640 <HOL>
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen
Query Match 5.3%; Score 23; DB 2; Length 1640;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 353 SGCFRHLDERECKCLLNKYQEG 375
DB 1542 SGCFRHLDERECKCLLNKYQEG 1564
RESULT 10
NR80B
pancreatic ribonuclease (EC 3.1.27.5) - American bison (tentative sequence)
N:Alternate names: RNase 1; RNase A
C:Species: Bison bison (American bison)
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 08-Dec-1994
C:Accession: A91771; A90270; A94696; A00804
R:Muskiet, F.A.J.; Welling, G.W.; Beintema, J.J.
Int. J. Pept. Protein Res. 8, 345-348, 1976
A:Reference number: A91771; MUID:76259396; PMID:955781
A:Accession: A91771
A:Molecule type: protein
A:Residues: 1-124 <MUS>
A:Note: comparison of peptide compositions and partial sequence determination revealed that the bison ribonuclease is identical to the bovine ribonuclease.
R:Stewart, G.R.; Stevenson, K.J.
Biochem. J. 135, 427-441, 1973
A:Title: The isolation and partial characterization of ribonuclease A from Bison bison.
A:Reference number: A90270; MUID:74081066; PMID:4772270
A:Accession: A90270

A:Molecule type: protein
A:Residues: 1-60,62-124 <STE>
A:Note: peptide compositions were compared with the bovine sequence
R:Barnard, E.A.; Cohen, M.S.; Gold, M.H.; Kim, J.K.
Nature 240, 395-398, 1972
A:Title: Evolution of ribonuclease in relation to polypeptide folding mechanisms.
A:Reference number: A94696; MUID:73045781; PMID:4564316
A:Accession: A94696
A:Molecule type: protein
A:Residues: 1-15,'E',17,'E',19-20,'XTA',24-30 <BAR>
C:Superfamily: pancreatic ribonuclease
C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas
F:12,41,119/Active site: His, Lys, His #status predicted
F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted
F:34/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 3.5%; Score 15; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 9.2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 KETAAAKFERQHMD 33
DB 1 KETAAAKFERQHMD 15
RESULT 11
NRANE
pancreatic ribonuclease (EC 3.1.27.5) - eland
N:Alternate names: RNase 1; RNase A
C:Species: Tragelaphus oryx, Taurotragus oryx (eland)
C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 31-Dec-1993
C:Accession: A00806
R:Russchen, F.; de Vriese, G.; Gaastra, W.; Beintema, J.J.
Biochim. Biophys. Acta 427, 719-726, 1976
A:Title: Studies on the covalent structure of eland pancreatic ribonuclease.
A:Reference number: A00806; MUID:76184786; PMID:1288225
A:Accession: A00806
A:Molecule type: protein
A:Residues: 1-124 <RUS>
C:Superfamily: pancreatic ribonuclease
C:Keywords: hydrolase; nucleic acid digestion; pancreas
F:12,41,119/Active site: His, Lys, His #status predicted
F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted
Query Match 3.5%; Score 15; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 9.2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 KETAAAKFERQHMD 33
DB 1 KETAAAKFERQHMD 15
RESULT 12
S08547
pancreatic ribonuclease (EC 3.1.27.5) - nilgai
C:Species: Boselaphus tragocamelus (nilgai)
C:Date: 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 05-Jun-1998
C:Accession: S08547
R:Beintema, J.J.
Biochim. Biophys. Acta 621, 89-103, 1980
A:Title: Primary structures of pancreatic ribonucleases from Bovidae. Impala, Thomson's gazelle, and nilgai.
A:Reference number: S07141; MUID:80109825; PMID:7353035
A:Accession: S08547
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-124 <BEI>
C:Superfamily: pancreatic ribonuclease
C:Keywords: hydrolase; nucleic acid digestion; pancreas
F:12,41,119/Active site: His, Lys, His #status predicted
F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted
Query Match 3.5%; Score 15; DB 2; Length 124;

Best Local Similarity 100.0%; Pred. No. 9.2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAKPERQHMS 33
|||||
Db 1 KETAAKPERQHMS 15

RESULT 13
JC5560
pancreatic ribonuclease (EC 3.1.27.5) A - Aspergillus niger var. macrosporus
C:Species: Aspergillus niger var. macrosporus
C:Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 07-May-1999
C:Accession: JC5560
R:Takahashi, K.
Biosci. Biotechnol. Biochem. 61, 381-383, 1997
A:Title: The specificity of peptide bond cleavage of acid proteinase A from Aspergillus
A:Reference number: JC5560; MUID:97212043; PMID:9058983
A:Accession: JC5560
A:Molecule type: protein
A:Residues: 1-124 <TAK>
C:Superfamily: pancreatic ribonuclease
C:Keywords: hydrolase

Query Match 3.5%; Score 15; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 9.2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAKPERQHMS 33
|||||
Db 1 KETAAKPERQHMS 15

RESULT 14
NR80
pancreatic ribonuclease (EC 3.1.27.5) precursor [validated] - bovine
N:Alternate names: ribonuclease I; ribonuclease A
C:Species: Bos primigenius taurus (cattle)
C:Date: 24-Apr-1984 #sequence_revision 05-Aug-1994 #text_change 15-Sep-2000
C:Accession: S00897; I45998; A32471; A00804; A92018; A61293; S53757; S05528
R:Carlsone, A.; Confalone, E.; Palmieri, M.; Libonati, M.; Furia, A.
Nucleic Acids Res. 16, 5491-5502, 1988
A:Title: Structure of the bovine pancreatic ribonuclease gene: the unique intervening se
A:Reference number: S00897; MUID:88262557; PMID:2838818
A:Accession: S00897
A:Molecule type: DNA
A:Residues: 1-150 <CAR>
A:Cross-references: EMBL:X07283; NID:g671; PIDN:CAA30263.1; PID:g672
R:Vasanthan, N.; Filipula, D.
Gene 76, 53-60, 1989
A:Title: Expression of bovine pancreatic ribonuclease A coded by a synthetic gene in Bac
A:Reference number: I45998; MUID:89306659; PMID:2501158
A:Accession: I45998
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 'M', 27-150 <VAS>
A:Cross-references: EMBL:X15802; NID:g93; PIDN:CAA33801.1; PID:g94
A:Note: recombinant gene expressed in E. coli
R:Robertson, A.D.; Purisima, E.O.; Eastman, M.A.; Scheraga, H.A.
Biochemistry 28, 5930-5938, 1989
A:Title: Proton NMR assignments and regular backbone structure of bovine pancreatic ribo
A:Reference number: A32471; MUID:89375325; PMID:2775743
A:Accession: A32471
A:Molecule type: protein
A:Residues: 27-150 <ROB>
R:Smyth, D.G.; Stein, W.H.; Moore, S.
J. Biol. Chem. 238, 227-234, 1963
A:Title: The sequence of amino acid residues in bovine pancreatic ribonuclease: revision
A:Reference number: A92016
A:Accession: A00804
A:Molecule type: protein
A:Residues: 27-150 <SMY>
A:Note: disulfide bonds were determined

R:Plummer Jr., T.H.; Hirs, C.H.W.
J. Biol. Chem. 239, 2530-2538, 1964
A:Title: On the structure of bovine pancreatic ribonuclease B. Isolation of a glycope
A:Reference number: A92018
A:Accession: A92018
A:Molecule type: protein
A:Residues: 27-150 <PLU>
R:Kumagai, H.; Yoshihara, K.; Umemoto, M.; Igarashi, K.; Hirose, S.; Ohgi, K.; Irie,
J. Biochem. 93, 865-874, 1983
A:Title: Studies on salivary gland ribonucleases. III. Purification and properties of
A:Reference number: A61293; MUID:83265672; PMID:6874668
A:Accession: A61293
A:Molecule type: protein
A:Residues: 27-40, 'XX', 43, 'X', 45-46 <KUM>
R:Yang, H.J.; Tsou, C.L.
Biochem. J. 305, 379-384, 1995
A:Title: Inactivation during denaturation of ribonuclease A by guanidinium chloride
A:Reference number: S53757; MUID:95134214; PMID:7832749
A:Accession: S53757
A:Molecule type: protein
A:Residues: 27-53; 60-63; 69-73 <YAN>
R:Birdsall, D.L.; McPherson, A.
submitted to the Brookhaven Protein Data Bank, August 1992
A:Reference number: A51385; PDB:1RTB
A:Contents: annotation; X-ray crystallography, 2.5 angstroms, residues 27-150
R:Birdsall, D.L.; McPherson, A.
J. Biol. Chem. 267, 22230-22236, 1992
A:Title: Crystal structure disposition of thymidylc acid tetramer in complex with ri
A:Reference number: A44321; MUID:93054504; PMID:1429575
A:Contents: annotation; X-ray crystallography, 2.5 angstroms
R:Williams, R.L.; Greene, S.M.; McPherson, A.
submitted to the Brookhaven Protein Data Bank, September 1987
A:Reference number: A50330; PDB:1RBB
A:Contents: annotation; X-ray crystallography, 2.5 angstroms, residues 27-150
R:Wlodawer, A.
submitted to the Brookhaven Protein Data Bank, April 1985
A:Reference number: A50927; PDB:5RSA
A:Contents: annotation; X-ray and neutron crystallography, 2.0 angstroms, residues 27
R:Wlodawer, A.; Bott, R.; Sjolín, L.
J. Biol. Chem. 257, 1325-1332, 1982
A:Title: The refined crystal structure of ribonuclease A at 2.0 angstrom resolution.
A:Reference number: A92350; MUID:82120062; PMID:6276380
A:Contents: annotation; X-ray crystallography, 2.0 angstroms
R:Howlin, B.; Moss, D.S.; Harris, G.W.; Palmer, R.A.
submitted to the Brookhaven Protein Data Bank, October 1991
A:Reference number: A50626; PDB:3RN3
A:Contents: annotation; X-ray crystallography, 1.45 angstroms, residues 27-150
R:Carlisle, C.H.; Palmer, R.A.; Mazumdar, S.K.; Gorinsky, B.A.; Yeates, D.G.R.
J. Mol. Biol. 85, 1-18, 1974
A:Title: The structure of ribonuclease at 2.5 angstrom resolution.
A:Reference number: A92949; MUID:74253324; PMID:4835726
A:Contents: annotation; X-ray crystallography, 2.5 angstroms
R:Wyckoff, H.W.; Tsernoglou, D.; Hanson, A.W.; Knox, J.R.; Lee, B.; Richards, F.M.
J. Biol. Chem. 245, 305-328, 1970
A:Title: The three-dimensional structure of ribonuclease-S. Interpretation of an elec
A:Reference number: A92059; MUID:70092235; PMID:5460889
A:Contents: annotation; X-ray crystallography, 2.0 angstroms
R:Santoro, J.; Gonzalez, C.; Bruix, M.; Neira, J.L.; Nieto, J.L.; Herranz, J.; Rico,
submitted to the Brookhaven Protein Data Bank, November 1992
A:Reference number: A52045; PDB:2AAS
A:Contents: annotation; conformation by (1)H-NMR, residues 27-150
R:Rico, M.; Bruix, M.; Santoro, J.; Gonzalez, C.; Neira, J.L.; Nieto, J.L.; Herranz,
Eur. J. Biochem. 183, 623-638, 1989
A:Title: Sequential (1)H-NMR assignment and solution structure of bovine pancreatic r
A:Reference number: S05528; MUID:89377830; PMID:2776756
A:Contents: annotation; confirmation by (1)H-NMR
R:Shall, S.; Barnard, E.A.
J. Mol. Biol. 41, 237-251, 1969
A:Title: Heavy atom-labelled derivatives of bovine pancreatic ribonuclease. I. Specif
A:Reference number: A92946; MUID:69260123; PMID:5801478
A:Contents: annotation; active site
R:Heinrikson, R.L.; Stein, W.H.; Crestfield, A.M.; Moore, S.
J. Biol. Chem. 240, 2921-2934, 1965

A:Title: The reactivities of the histidine residues at the active site of ribonuclease d
A:Reference number: A92020
A:Contents: annotation; active site
C:Function:
A:Description: ribonuclease endonucleolytically hydrolyzes RNA to produce 3'-phosphomono
A:Pathway: ribonucleic acid digestion
A:Note: this enzyme can also destabilize or unwind the DNA helix by complexing with sing
rginine residues of the enzyme and the phosphate groups of the nucleotides
C:Superfamily: pancreatic ribonuclease
C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-150/Product: pancreatic ribonuclease #status experimental <MAT>
F:38,67,145/Active site: His, Lys, His #status experimental
F:52-110,66-121,84-136,91-98/Disulfide bonds: #status experimental
F:60/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 3.5% Score 15; DB 1; Length 150;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAKFERQHMD 33
|||||
DB 27 KETAAKFERQHMD 41

RESULT 15
A05004
pancreatic ribonuclease (EC 3.1.27.5) - white-tailed deer (fragment)
N:Alternate names: RNase 1; RNase A
C:Species: Odocoileus virginianus (white-tailed deer)
C:Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 28-Apr-1993
C:Accession: A05004
R:Barnard, E.A.; Cohen, M.S.; Gold, M.H.; Kim, J.K.
Nature 240, 395-398, 1972
A:Title: Evolution of ribonuclease in relation to polypeptide folding mechanisms.
A:Reference number: A94696; MUID:73045781; PMID:4564316
A:Accession: A05004
A:Molecule type: protein
A:Residues: 1-30 <BAR>
A:Note: Ser-3 was also found
C:Superfamily: pancreatic ribonuclease
C:Keywords: hydrolase; nucleic acid digestion; pancreas

Query Match 3.2% Score 14; DB 2; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.6e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAKFERQHMD 32
|||||
DB 1 KETAAKFERQHMD 14

RESULT 16
NRGN
pancreatic ribonuclease (EC 3.1.27.5) - brindled gnu
N:Alternate names: RNase 1; RNase A
C:Species: Connochaetes taurinus (brindled gnu, white-bearded gnu, blue wildebeest)
C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 04-Oct-1996
C:Accession: A00808
R:Groen, G.; Welling, G.W.; Beintema, J.J.
FEBS Lett. 60, 300-304, 1975
A:Title: The amino acid sequence of gnu pancreatic ribonuclease.
A:Reference number: A00808; MUID:76210818; PMID:1227969
A:Accession: A00808
A:Molecule type: protein
A:Residues: 1-124 <GRO>
C:Superfamily: pancreatic ribonuclease
C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas
F:12,41,119/Active site: His, Lys, His #status predicted
F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted
F:34/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.8% Score 12; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00077;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 0.00077;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAAKFERQHMD 33
|||||
DB 4 AAAKFERQHMD 15

RESULT 17
NRSH
pancreatic ribonuclease (EC 3.1.27.5) - sheep
N:Alternate names: RNase 1; RNase A
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 04-Oct-1996
C:Accession: A91406; A92145; A00809
R:Welling, G.W.; Scheffer, A.J.; Beintema, J.J.
FEBS Lett. 41, 58-61, 1974
A:Title: The primary structure of goat and sheep pancreatic ribonucleases.
A:Reference number: A91406; MUID:74309062; PMID:4855010
A:Accession: A91406
A:Molecule type: protein
A:Residues: 1-124 <WEL>
R:Kobayashi, R.; Hirs, C.H.W.
J. Biol. Chem. 248, 7833-7837, 1973
A:Title: The amino acid sequence of ovine pancreatic ribonuclease A.
A:Reference number: A92145; MUID:74025520; PMID:4356260
A:Accession: A92145
A:Molecule type: protein
A:Residues: 1-48, 'O', 50-102, 'O', 104-124 <KOB>
C:Superfamily: pancreatic ribonuclease
C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas
F:12,41,119/Active site: His, Lys, His #status predicted
F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted
F:34/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.8% Score 12; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00077;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAAKFERQHMD 33
|||||
DB 4 AAAKFERQHMD 15

RESULT 18
NRGT
pancreatic ribonuclease (EC 3.1.27.5) - goat
N:Alternate names: RNase 1; RNase A
C:Species: Capra aegagrus hircus (domestic goat)
C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 04-Oct-1996
C:Accession: A00810
R:Welling, G.W.; Scheffer, A.J.; Beintema, J.J.
FEBS Lett. 41, 58-61, 1974
A:Title: The primary structure of goat and sheep pancreatic ribonucleases.
A:Reference number: A91406; MUID:74309062; PMID:4855010
A:Accession: A00810
A:Molecule type: protein
A:Residues: 1-124 <WEL>
C:Superfamily: pancreatic ribonuclease
C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas
F:12,41,119/Active site: His, Lys, His #status predicted
F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted
F:34/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.8% Score 12; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00077;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAAKFERQHMD 33
|||||
DB 4 AAAKFERQHMD 15

RESULT 19

NRPRH
pancreatic ribonuclease (EC 3.1.27.5) - pronghorn (tentative sequence)
N:Alternate names: RNase 1; RNase A
C:Species: Antilocapra americana (pronghorn)
C>Date: 28-Feb-1981 #sequence_revision 28-Feb-1981 #text_change 31-Mar-2000
C:Accession: A00813
R:Beintema, J.J.; Gastra, W.; Munnikma, J.
J. Mol. Evol. 13, 305-316, 1979
A:Title: Primary structure of pronghorn pancreatic ribonuclease: close relationship between
A:Reference number: A00813; MUID:80075014; PMID:513141
A:Accession: A00813
A:Molecule type: protein
A:Residues: 1-124 <BEI>
C:Superfamily: pancreatic ribonuclease
C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas
F:12,41,119/Active site: His, Lys, His #status predicted
F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted
F:34/Blinding site: carbohydrate (Asn) (covalent) (partial) #status experimental

Query Match 2.8%; Score 12; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00077;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KETAAAKFERQH 30

|||||

DB 1 KETAAAKFERQH 12

RESULT 20

S07141
pancreatic ribonuclease (EC 3.1.27.5) - impala
C:Species: Aepyceros melampus (impala)
C>Date: 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 14-Nov-1997
C:Accession: S07141
R:Beintema, J.J.
R:Belntema, J.J.
Biochim. Biophys. Acta 621, 89-103, 1980

A:Title: Primary structures of pancreatic ribonucleases from Bovidae. Impala, Thomson's
A:Reference number: S07141; MUID:80109825; PMID:7353035
A:Accession: S07141
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-124 <BEI>

C:Superfamily: pancreatic ribonuclease
C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas
F:12,41,119/Active site: His, Lys, His #status predicted
F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

Query Match 2.8%; Score 12; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00077;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAACFERQHMS 33

|||||

DB 4 AAACFERQHMS 15

RESULT 21

A47498
seminal ribonuclease (EC 3.1.27.5) - synthetic
C:Species: synthetic
A:Note: Bos primigenius taurus (cattle) gene engineered and expressed in Escherichia coli
C>Date: 16-Feb-1994 #sequence_revision 30-Nov-1995 #text_change 30-Nov-1995
C:Accession: A47498
R:Kim, J.S.; Raines, R.T.
J. Biol. Chem. 268, 17392-17396, 1993
A:Title: Bovine seminal ribonuclease produced from a synthetic gene.
A:Reference number: A47498; MUID:93352527; PMID:7688724
A:Accession: A47498
A:Molecule type: nucleic acid
A:Residues: 1-125 <KIM>

A:Note: sequence extracted from NCBI backbone (NCBIN:136293, NCBIIP:136294)
A:Note: part of this sequence, including the amino end of the mature protein, was determined

C:Keywords: hydrolase

F:27-85,41-96,59-111,66-73/Disulfide bonds: #status predicted
F:32,33/Disulfide bonds: Interchain #status predicted

Query Match 2.8%; Score 12; DB 4; Length 125;
Best Local Similarity 100.0%; Pred. No. 0.00078;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAACFERQHMS 33

|||||

DB 5 AAACFERQHMS 16

RESULT 22

NRBOS

seminal ribonuclease (EC 3.1.27.5) precursor - bovine

C:Species: Bos primigenius taurus (cattle)

C>Date: 24-Apr-1984 #sequence_revision 30-Jun-1991 #text_change 18-Jun-1999

R:Preuss, K.D.; Wagner, S.; Freudenstein, J.; Scheit, K.H.

Nucleic Acids Res. 18, 1057, 1990

A:Title: Cloning of cDNA encoding the complete precursor for bovine seminal ribonuclease

A:Reference number: S08392; MUID:90192098; PMID:2315023

A:Accession: S08392

A:Molecule type: mRNA

A:Residues: 1-150 <PRE>

A:Cross-references: EMBL:X51337; NID:g748; PIDN:CAA35716.1; PID:g749

R:Suzuki, H.; Parente, A.; Farina, B.; Greco, L.; la Montagna, R.; Leone, E.

Biol. Chem. Hoppe-Seyler 368, 1305-1312, 1987

A:Title: Complete amino-acid sequence of bovine seminal ribonuclease, a dimeric protein

A:Reference number: S00131; MUID:88106995; PMID:3426801

A:Accession: S00131

A:Molecule type: protein

A:Residues: 27-150 <SUZ>

A:Note: this is the final paper in a series

R:Ramakrishna, T.; Vijayarangam, D.; Sitaram, N.; Pandit, M.W.; Bhargava, P.M.

Biochem. Int. 26, 125-133, 1992

A:Title: Apparent specificity of bovine seminal ribonucleases can depend on the conditions

A:Reference number: A45654; MUID:92313367; PMID:1377468

A:Accession: A45654

A:Molecule type: protein

A:Residues: 27-42, 'X', 44-47 <RAM>

A:Experimental source: seminal plasma

A:Note: sequence extracted from NCBI backbone (NCBIP:107298)

R:Palmeri, M.; Carsana, A.; Furia, A.; Libonati, M.

Eur. J. Biochem. 152, 275-277, 1985

A:Title: Sequence analysis of a cloned cDNA coding for bovine seminal ribonuclease.

A:Reference number: A23094; MUID:86030265; PMID:3840434

A:Accession: A23094

A:Molecule type: mRNA

A:Residues: 73-150 <PAL>

A:Cross-references: GB:X03029; NID:g752; PIDN:CAA26832.1; PID:g753

R:Krietsch, W.K.G.; Simm, F.C.; Hertenberger, B.; Kuntz, G.W.K.; Wachter, E.

Anal. Biochem. 128, 213-216, 1983

A:Title: Isolation of bovine seminal ribonuclease by affinity chromatography.

A:Reference number: A30432; MUID:83202042; PMID:6846794

A:Contents: annotation; amidation state of residue 43

R:di Donato, A.; d'Alessio, G.

Biochem. Biophys. Res. Commun. 55, 919-928, 1973

A:Title: Interchain disulfide bridges in ribonuclease BS-1.

A:Reference number: A30433; MUID:74049879; PMID:4761089

A:Contents: annotation; interchain disulfide bonds

R:di Donato, A.; d'Alessio, G.

Biochim. Biophys. Acta 579, 303-313, 1979

A:Title: Intrachain disulfide bridges of bovine seminal ribonuclease.

A:Reference number: A30434; MUID:80130655; PMID:534646

A:Contents: annotation; intrachain disulfide bonds

R:Capasso, S.; Giordano, F.; Mattia, C.A.; Mazzarella, L.; Zagari, A.

Biopolymers 22, 327-332, 1983

A:Title: Refinement of the structure of bovine seminal ribonuclease.

A:Reference number: A30435; MUID:84179398; PMID:6673761

A:Contents: annotation; X-ray crystallography, 2.5 angstroms

C:Comment: This enzyme hydrolyzes both single- and double-stranded RNA.

C:Superfamily: pancreatic ribonuclease
 C:Keywords: homodimer; hydrolase
 F:1-26/Domain: signal sequence #status predicted <SIG>
 F:27-150/Product: seminal ribonuclease #status experimental <MAT>
 F:38,67,145/Active site: His, Lys, His #status predicted
 F:52-110,66-121,84-136,91-99/Disulfide bonds: #status experimental
 F:57,58/Disulfide bonds: interchain #status experimental

Query Match 2.8%; Score 12; DB 1; Length 150;
 Best Local Similarity 100.0%; Pred. No. 0.0091;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 22 AAAXFEROHMDS 33
 |||||
 Db 30 AAAXFERQHMDs 41

RESULT 23

NRDEF
 pancreatic ribonuclease (EC 3.1.27.5) - red deer (tentative sequence)
 N:Alternate names: RNase 1; RNase A
 C:Species: Cervus elaphus (red deer)
 C:Date: 24-Apr-1984 #sequence_revision 31-Dec-1991 #text_change 03-Feb-1994
 C:Accession: B90613; A00812
 R:Zwijers, H.; Schaffer, A.J.; Beintema, J.J.
 Eur. J. Biochem. 36, 569-574, 1973
 A:Title: Amino-acid sequences of red-deer and roe-deer pancreatic ribonucleases.
 A:Reference number: A91207; MUID:73253236; PMID:4738402
 A:Contents: annotation
 A:Note: the sequence has been revised in reference A90613
 R:Oosterhuis, S.; Wellington, G.W.; Gastra, W.; Beintema, J.J.
 Biochim. Biophys. Acta 490, 523-529, 1977
 A:Title: Reinvestigation of the primary structures of red deer and roe deer pancreatic
 A:Reference number: A90613; MUID:77112628; PMID:836889
 A:Accession: B90613
 A:Molecule type: protein
 A:Residues: 1-124 <OOS>
 A:Note: the authors reinvestigated the sequences of residues 1-25

C:Superfamily: pancreatic ribonuclease
 C:Keywords: hydrolase; nucleic acid digestion; pancreas
 F:12,41,119/Active site: His, Lys, His #status predicted
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

Query Match 2.6%; Score 11; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.0073;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 22 AAAXFERQHMD 32
 |||||
 Db 4 AAAXFERQHMD 14

RESULT 24

NRDEO
 pancreatic ribonuclease (EC 3.1.27.5) - roe deer (tentative sequence)
 N:Alternate names: RNase 1; RNase A
 C:Species: Capreolus capreolus (roe deer)
 C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 03-Feb-1994
 C:Accession: A90613; A00812
 R:Zwijers, H.; Schaffer, A.J.; Beintema, J.J.
 Eur. J. Biochem. 36, 569-574, 1973
 A:Title: Amino-acid sequences of red-deer and roe-deer pancreatic ribonucleases.
 A:Reference number: A91207; MUID:73253236; PMID:4738402
 A:Contents: annotation
 A:Note: the sequence has been revised in reference A90613
 R:Oosterhuis, S.; Wellington, G.W.; Gastra, W.; Beintema, J.J.
 Biochim. Biophys. Acta 490, 523-529, 1977
 A:Title: Reinvestigation of the primary structures of red deer and roe deer pancreatic
 A:Reference number: A90613; MUID:77112628; PMID:836889
 A:Accession: A90613
 A:Molecule type: protein
 A:Residues: 1-124 <OOS>
 A:Note: the authors reinvestigated the sequences of residues 1-25 and 99

C:Superfamily: pancreatic ribonuclease
 C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas
 F:12,41,119/Active site: His, Lys, His #status predicted
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted
 F:34/Binding site: carbohydrate (Asn) (covalent) (partial) #status experimental

Query Match 2.6%; Score 11; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.0073;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 22 AAAXFERQHMD 32
 |||||
 Db 4 AAAXFERQHMD 14

RESULT 25

NRDEF
 pancreatic ribonuclease (EC 3.1.27.5) - fallow deer (tentative sequence)
 N:Alternate names: RNase 1; RNase A
 C:Species: Dama dama (fallow deer)
 C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 31-Dec-1993
 C:Accession: A91418; A00812
 R:Leijenaar-van den Berg, G.; Beintema, J.J.
 FEBS Lett. 56, 101-107, 1975
 A:Title: The amino acid sequences of reindeer, moose and fallow deer pancreatic ribon
 A:Reference number: A91418; MUID:76003215; PMID:1157925
 A:Accession: A91418
 A:Molecule type: protein
 A:Residues: 1-124 <LEI>

C:Superfamily: pancreatic ribonuclease
 C:Keywords: hydrolase; nucleic acid digestion; pancreas
 F:12,41,119/Active site: His, Lys, His #status predicted
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

Query Match 2.6%; Score 11; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.0073;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 22 AAAXFERQHMD 32
 |||||
 Db 4 AAAXFERQHMD 14

RESULT 26

NRDEN
 pancreatic ribonuclease (EC 3.1.27.5) - reindeer (tentative sequence)
 N:Alternate names: RNase 1; RNase A
 C:Species: Rangifer tarandus (reindeer)
 C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 31-Dec-1993
 C:Accession: B91418; A00812
 R:Leijenaar-van den Berg, G.; Beintema, J.J.
 FEBS Lett. 56, 101-107, 1975
 A:Title: The amino acid sequences of reindeer, moose and fallow deer pancreatic ribon

A:Reference number: A91418; MUID:76003215; PMID:1157925
 A:Accession: B91418
 A:Molecule type: protein

C:Superfamily: pancreatic ribonuclease
 C:Keywords: hydrolase; nucleic acid digestion; pancreas
 F:12,41,119/Active site: His, Lys, His #status predicted
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

Query Match 2.6%; Score 11; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.0073;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 22 AAAXFERQHMD 32
 |||||
 Db 4 AAAXFERQHMD 14

RESULT 27

NREKN

pancreatic ribonuclease (EC 3.1.27.5) - European moose (tentative sequence)
 N:Alternate names: RNase 1; RNase A
 C:Species: Alces alces alces (European moose, elk)
 C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 31-Dec-1993
 C:Accession: C91418; A00812
 R:Leijenaar-van den Berg, G.; Beintema, J.J.
 FEBS Lett. 56, 101-107, 1975

A:Title: The amino acid sequences of reindeer, moose and fallow deer pancreatic ribonuclease
 A:Reference number: A91418; MUID:76003215; PMID:1157925

A:Accession: C91418
 A:Molecule type: protein
 A:Residues: 1-124 <LET>
 C:Superfamily: pancreatic ribonuclease
 C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas
 F:12,41,119/Active site: His, Lys, His #status predicted
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted
 F:34/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 2.6%; Score 11; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.0073;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAAKFERQHMD 32
 DB 4 AAAKFERQHMD 14
 |||||

RESULT 28
 NRCM

pancreatic ribonuclease (EC 3.1.27.5) - giraffe (tentative sequence)

N:Alternate names: RNase 1; RNase A
 C:Species: Giraffa camelopardalis (giraffe)
 C:Date: 24-Apr-1984 #sequence_revision 24-Apr-1984 #text_change 31-Mar-2000
 C:Accession: A94452; A91405; A00811

R:Gaastra, W.
 Ph.D. thesis, University of Groningen, Groningen, The Netherlands, 1975

A:Reference number: A94452
 A:Accession: A94452
 A:Molecule type: protein
 A:Residues: 1-124 <GAA>
 R:Gaastra, W.; Groen, G.W.; Beintema, J.J.
 FEBS Lett. 41, 227-232, 1974
 A:Title: The primary structure of giraffe pancreatic ribonuclease.
 A:Reference number: A91405; MUID:74309061; PMID:4855009

A:Accession: A91405
 A:Molecule type: protein
 A:Residues: 1-124 <GA2>
 C:Superfamily: pancreatic ribonuclease
 C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas
 F:12,41,119/Active site: His, Lys, His #status predicted
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted
 F:34/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 2.1%; Score 9; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.65;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAAKFERQH 30
 DB 4 AAAKFERQH 12
 |||||

RESULT 29
 NRCM

pancreatic ribonuclease (EC 3.1.27.5) - Arabian camel (tentative sequence)

N:Alternate names: RNase 1; RNase A
 C:Species: Camelus dromedarius (Arabian camel)
 C:Date: 24-Apr-1984 #sequence_revision 17-Mar-1987 #text_change 31-Dec-1993
 C:Accession: A00815; A90283

R:Beintema, J.J.
 FEBS Lett. 185, 115-120, 1985

A:Title: Mammalian ribonucleases. The absence of a glycosylated Asn-Pro-Thr sequence in
 A:Reference number: A91340; MUID:85204378; PMID:3922790

A:Accession: A00815
 A:Molecule type: protein
 A:Residues: 1-124 <BEI>
 R:Wellington, G.W.; Groen, G.; Beintema, J.J.
 Biochem. J. 147, 505-511, 1975
 A:Title: The amino acid sequence of dromedary pancreatic ribonuclease.
 A:Reference number: A90283; MUID:76039472; PMID:1167157

A:Accession: A90283
 A:Molecule type: protein
 A:Residues: 1-36, N', 38, 'X', 40-77, 'S', 79-102, 'Q', 104-124 <WEL>
 A:Note: this sequence has been revised in reference A91340
 C:Superfamily: pancreatic ribonuclease
 C:Keywords: hydrolase; nucleic acid digestion; pancreas
 F:12,41,119/Active site: His, Lys, His #status predicted
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

Query Match 2.1%; Score 9; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.65;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMD 33
 DB 7 KFERQHMD 15
 |||||

RESULT 30
 NRCM

pancreatic ribonuclease (EC 3.1.27.5), minor form - Arabian camel (tentative sequence)

N:Alternate names: RNase 1; RNase A
 C:Species: Camelus dromedarius (Arabian camel)
 C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 31-Dec-1993
 C:Accession: A90229; A00815

R:Wellington, G.W.; Mulder, H.; Beintema, J.J.
 Biochem. Genet. 14, 309-317, 1976

A:Title: Allelic polymorphism in Arabian camel ribonuclease and the amino acid sequence
 A:Reference number: A90229; MUID:76277807; PMID:962846

A:Accession: A90229
 A:Molecule type: protein
 A:Residues: 1-124 <WEL>
 C:Superfamily: pancreatic ribonuclease
 C:Keywords: hydrolase; nucleic acid digestion; pancreas
 F:12,41,119/Active site: His, Lys, His #status predicted
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

Query Match 2.1%; Score 9; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.65;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMD 33
 DB 7 KFERQHMD 15
 |||||

RESULT 31
 NRCM

pancreatic ribonuclease (EC 3.1.27.5) - Bactrian camel (tentative sequence)

N:Alternate names: RNase 1; RNase A
 C:Species: Camelus bactrianus (Bactrian camel)
 C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 31-Dec-1993
 C:Accession: B90229; A00815

R:Wellington, G.W.; Mulder, H.; Beintema, J.J.
 Biochem. Genet. 14, 309-317, 1976

A:Title: Allelic polymorphism in Arabian camel ribonuclease and the amino acid sequence
 A:Reference number: A90229; MUID:76277807; PMID:962846

A:Accession: B90229
 A:Molecule type: protein
 A:Residues: 1-124 <WEL>
 C:Superfamily: pancreatic ribonuclease
 C:Keywords: hydrolase; nucleic acid digestion; pancreas
 F:12,41,119/Active site: His, Lys, His #status predicted
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

Query Match 2.1%; Score 9; DB 1; Length 124;

Best Local Similarity 100.0%; Pred. No. 0.65; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0;

QY 25 KFERQHMD 33
| | | | | | | |
DB 7 KFERQHMD 15

RESULT 32

NRQ

pancreatic ribonuclease (EC 3.1.27.5) - muskrat (tentative sequence)

N:Alternate names: RNase 1; RNase A

C:Species: Ondatra zibethicus (muskrat)

C:Date: 24-Apr-1984 #sequence_revision 24-Apr-1984 #text_change 31-Mar-2000

C:Accession: A00828

R:van Dijk, H.; Sloots, B.; van den Berg, A.; Gastra, W.; Beintema, J.J.

I:nt. J. Pept. Protein Res. 8, 305-316, 1976

A:Title: The primary structure of muskrat pancreatic ribonuclease.

A:Reference number: A00828; MUID:76212474; PMID:1279085

A:Accession: A00828

A:Molecule type: protein

A:Residues: 1-124 <VAN>

A:Note: The placement of residues 32-33 and 75-76 is tentative

C:Superfamily: pancreatic ribonuclease

C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas

F:12.41.119/Active site: His, Lys, His #status predicted

F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

F:62/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 2.1%; Score 9; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.65;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMD 33
| | | | | | | |
DB 7 KFERQHMD 15

RESULT 33

NRHY

pancreatic ribonuclease (EC 3.1.27.5) - golden hamster (tentative sequence)

N:Alternate names: RNase 1; RNase A

C:Species: Mesocricetus auratus (golden hamster)

C:Date: 30-Nov-1980 #sequence_revision 30-Nov-1980 #text_change 31-Mar-2000

C:Accession: A00829

R:Jekel, P.A.; Sips, H.J.; Lenstra, J.A.; Beintema, J.J.

Biochimie 61, 827-839, 1979

A:Title: The amino acid sequence of hamster pancreatic ribonuclease.

A:Reference number: A00829; MUID:80088445; PMID:518928

A:Accession: A00829

A:Molecule type: protein

A:Residues: 1-124 <JEK>

C:Superfamily: pancreatic ribonuclease

C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas

F:12.41.119/Active site: His, Lys, His #status predicted

F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

F:34/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 2.1%; Score 9; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.65;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMD 33
| | | | | | | |
DB 7 KFERQHMD 15

RESULT 34

NRHO

pancreatic ribonuclease (EC 3.1.27.5) - horse

N:Alternate names: RNase 1; RNase A

C:Species: Equus caballus (domestic horse)

C:Date: 24-Apr-1984 #sequence_revision 17-Mar-1987 #text_change 31-Dec-1993

C:Accession: A91340; A00819
R:Beintema, J.J.
FEBS Lett. 185, 115-120, 1985

A:Title: Mammalian ribonucleases. The absence of a glycosylated Asn-Pro-Thr sequence

A:Reference number: A91340; MUID:85204378; PMID:3922790

A:Accession: A91340

A:Molecule type: protein

A:Residues: 1-128 <BEI>

R:Scheffer, A.J.; Beintema, J.J.

Eur. J. Biochem. 46, 221-233, 1974

A:Title: Horse pancreatic ribonuclease.

A:Reference number: A91222; MUID:74302367; PMID:4852291

A:Contents: annotation; tentative sequence

A:Note: this sequence has been revised

C:Superfamily: pancreatic ribonuclease

C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas

F:12.41.119/Active site: His, Lys, His #status predicted

F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

F:34/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 2.1%; Score 9; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.66;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMD 33
| | | | | | | |
DB 7 KFERQHMD 15

RESULT 35

NRQU

pancreatic ribonuclease (EC 3.1.27.5) - nutria (tentative sequence)

N:Alternate names: RNase 1; RNase A

C:Species: Myocastor coypus (nutria, coypu)

C:Date: 24-Apr-1984 #sequence_revision 30-Sep-1988 #text_change 31-Mar-2000

C:Accession: A00822

R:van den Berg, A.; van den Hende-Timmer, L.; Beintema, J.J.

Biochim. Biophys. Acta 453, 400-409, 1976

A:Title: Isolation, properties and primary structure of coypu and chinchilla pancreat

A:Reference number: A90612; MUID:77065676; PMID:999896

A:Accession: A00822

A:Molecule type: protein

A:Residues: 1-128 <VAN>

C:Superfamily: pancreatic ribonuclease

C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas

F:12.41.119/Active site: His, Lys, His #status predicted

F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted

F:34/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 2.1%; Score 9; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.66;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMD 33
| | | | | | | |
DB 7 KFERQHMD 15

RESULT 36

NRQ

pancreatic ribonuclease (EC 3.1.27.5) - crested porcupine

C:Species: Hystrix cristata (crested porcupine)

C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 30-Sep-1993

C:Accession: A00823

R:Beintema, J.J.; Knol, G.; Martena, B.

Biochim. Biophys. Acta 705, 102-110, 1982

A:Title: The primary structures of pancreatic ribonucleases from African porcupine an

A:Reference number: A90644; MUID:83000399; PMID:7115727

A:Accession: A00823

A:Molecule type: protein

A:Residues: 1-128 <BEI>

A:Note: residues 67-78 were positioned primarily by homology with other ribonucleases

A:Note: 98-Arg.was also found

C:Superfamily: pancreatic ribonuclease
 C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas
 F:12.41.119/Active site: His, Lys, His #status predicted
 F:26-84.40-95.58-110.65-72/Disulfide bonds: #status predicted
 F:34/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 2.1%; Score 9; DB 1; Length 128;
 Best Local Similarity 100.0%; Pred. No. 0.66;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMD 33
 |||||
 Db 7 KFERQHMD 15

RESULT 37

S22808
 pancreatic ribonuclease (EC 3.1.27.5) - long-tailed hamster (fragment)
 C:Species: Cricetulus longicaudatus (long-tailed hamster)
 C:Date: 22-Nov-1993 #sequence_revision 09-Aug-1997 #text_change 22-Jun-1999
 C:Accession: S22808

R:Haugg, M.; Schein, C.H.
 Nucleic Acids Res. 20, 612, 1992
 A:Title: The DNA sequences of the human and hamster secretory ribonucleases determined
 A:Reference number: S22808; MUID:92158677; PMID:1741299

A:Accession: S22808
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA

A:Residues: 1-17, 'X' 19-130 <HAD>
 A:Cross-references: EMBL:X62945; NID:g49519; PIDN:CAA44717.1; PID:g938047
 A:Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1991
 A:Note: we have shown the unidentified residue 18 as His forming a conserved active site

C:Superfamily: pancreatic ribonuclease
 C:Keywords: glycoprotein; hydrolase
 F:18.47.125/Active site: His, Lys, His #status predicted
 F:32-90.46-101.64-116.71-78/Disulfide bonds: #status predicted

Query Match 2.1%; Score 9; DB 2; Length 130;
 Best Local Similarity 100.0%; Pred. No. 0.67;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMD 33
 |||||
 Db 13 KFERQHMD 21

RESULT 38

RSBPL5

Integrase - Staphylococcus phage L54a

C:Species: Staphylococcus phage L54a

A:Note: host Staphylococcus aureus

C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 23-Jul-1999

C:Accession: A33855

R:Ye, Z.H.; Lee, C.Y.

J. Bacteriol. 171, 4146-4153, 1989

A:Title: Nucleotide sequence and genetic characterization of staphylococcal bacteriophage

A:Reference number: A33855; MUID:89327124; PMID:2526804

A:Accession: A33855

A:Molecule type: DNA

A:Residues: 1-59 <YE2>

A:Cross-references: GB:M27965; NID:g215096; PIDN:AAA98160.1; PID:g455285

C:Gene: int

C:Superfamily: phage L54a integrase

C:Keywords: DNA binding; DNA integration

Query Match 1.9%; Score 8; DB 1; Length 59;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 YIKESVEN 167
 |||||
 Db 12 YIKESVEN 19

RESULT 39

C64698

probable histidine-rich metal-binding protein - Helicobacter pylori

C:Species: Helicobacter pylori

A:Variety: strains J99, 26695

C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999

C:Accession: C64698; C71821

R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKee,

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: C64698

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-60 <TOM>

A:Cross-references: GB:AE000643; GB:AE000511; NID:g2314598; PIDN:AAD08471.1; PID:g231

A:Experimental source: strain 26695

R:Alm, R.A.; Ling, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.

Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p

A:Reference number: A71800; MUID:99120557; PMID:9923682

A:Accession: C71821

A:Molecule type: DNA

A:Residues: 1-60 <ARN>

A:Cross-references: GB:AE001555; GB:AE001439; NID:g4155929; PIDN:AAD06898.1; PID:g415

A:Experimental source: strain J99

C:Genetics:

A:Gene: HP1427; jhp1320

Query Match 1.9%; Score 8; DB 2; Length 60;

Best Local Similarity 100.0%; Pred. No. 3.3;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHHSS 9

|||||

Db 28 HHHHHHSS 35

RESULT 40

NRWB

pancreatic ribonuclease (EC 3.1.27.5) - domestic water buffalo

N:Alternate names: RNase 1; RNase A

C:Species: Bubalus arnee bubalis (domestic water buffalo)

C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 01-Dec-1995

C:Accession: A00805; S08548

R:Sidik, A.; Martena, B.; Beintema, J.J.

Biochem. Genet. 17, 1151-1158, 1979

A:Title: Amino acid sequence differences in pancreatic ribonucleases from water buffa

A:Reference number: A00805; MUID:80153329; PMID:540006

A:Accession: A00805

A:Molecule type: protein

A:Residues: 1-124 <SID>

A:Experimental source: Italian river breed and Indonesian swamp breed

A:Note: the sequence from the Italian river breed is shown

A:Note: the sequence from Indonesian swamp breed differs from that shown in having 34

a polymorphic form that does

R:Beintema, J.J.

Biochim. Biophys. Acta 621, 89-103, 1980

A:Title: Primary structures of pancreatic ribonucleases from Bovidae. Impala, Thomson

A:Reference number: S07141; MUID:80109825; PMID:7353035

A:Accession: S08548

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-124 <BE1>

C:Superfamily: pancreatic ribonuclease

C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas

F:12.41.119/Active site: His, Lys, His #status predicted

F;26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted
F;34/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 1.9% Score 8; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 KETAAAKF 26
| | | | | | | |
Db 1 KETAAAKF 8

Search completed: May 12, 2003, 10:28:41
Job time : 53 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 10:20:38 ; Search time 14 Seconds
(without alignments)
1276.878 Million cell updates/sec

Title: US-10-057-531A-2

Perfect score: 431

Sequence: 1 MHHHHHHHSLVPRGSMKE.....TCECTKPSYPLFDGIFCSS 431

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	231	53.6	1726	1 MSP1_PLAFC
2	231	53.6	1726	1 MSP1_PLAFC
3	190	44.1	1701	1 MSP1_PLAFC
4	190	44.1	1701	1 MSP1_PLAFC
5	172	39.9	1682	1 MSP1_PLAFC
6	46	10.7	1630	1 MSP1_PLAFC
7	46	10.7	1639	1 MSP1_PLAFC
8	15	3.5	124	1 RNP_BOSTR
9	15	3.5	124	1 RNP_TRAOR
10	15	3.5	150	1 RNP_BOVIN
11	14	3.2	30	1 RNP_ODOVI
12	12	2.8	124	1 RNP_AEPME
13	12	2.8	124	1 RNP_ANTAM
14	12	2.8	124	1 RNP_CONTA
15	12	2.8	124	1 RNP_SHEEP
16	12	2.8	150	1 RNS_BOVIN
17	11	2.6	124	1 RNP_ALCAA
18	11	2.6	124	1 RNP_AXIPR
19	11	2.6	124	1 RNP_CAPCA
20	11	2.6	124	1 RNP_CEREL
21	11	2.6	124	1 RNP_DAMDA
22	11	2.6	124	1 RNP_RANTA
23	9	2.1	124	1 RNP_CAMDR
24	9	2.1	124	1 RNP_GIRCA
25	9	2.1	124	1 RNP_MESAU
26	9	2.1	124	1 RNP_ONDZI
27	9	2.1	128	1 RNP_HORSE
28	9	2.1	128	1 RNP_HYSCR
29	9	2.1	128	1 RNP_MYOCO
30	9	2.1	130	1 RNP_CRILQ
31	9	2.1	148	1 RNP_CLEGL
32	9	2.1	148	1 RNP_GERNI
33	9	2.1	148	1 RNP_TATKG

34	9	2.1	149	1 RNP_ACOCA
35	9	2.1	149	1 RNP_AKOJE
36	9	2.1	149	1 RNP_MUSPA
37	9	2.1	149	1 RNP_URARU
38	8	1.9	59	1 HPN_HELPY
39	8	1.9	59	1 VXIS_BPL34
40	8	1.9	97	1 EOTA_MOUSE
41	8	1.9	97	1 EOTA_RAT
42	8	1.9	124	1 RNP_BUBBU
43	8	1.9	141	1 RNR_GIRCA
44	8	1.9	231	1 YE93_METJA
45	8	1.9	582	1 SEM2_HUMAN
46	8	1.9	583	1 PGM1_MAIZE
47	8	1.9	583	1 PGM2_MAIZE
48	8	1.9	717	1 UBPB_YEAST
49	8	1.9	972	1 TOP1_DROME
50	8	1.9	972	1 ROM2_YEAST
51	7	1.6	182	1 YN66_HALNI
52	7	1.6	196	1 MOBA_STACA
53	7	1.6	216	1 FUCA_HAEIN
54	7	1.6	216	1 HAN1_MOUSE
55	7	1.6	216	1 HAN1_RAT
56	7	1.6	232	1 HXB4_XENLA
57	7	1.6	260	1 YH20_PASMO
58	7	1.6	279	1 BMRR_BACSU
59	7	1.6	314	1 YK83_MYCTU
60	7	1.6	328	1 YQ51_CAEEL
61	7	1.6	331	1 SUCA_SCHPO
62	7	1.6	342	1 RFL1_THEMA
63	7	1.6	374	1 YMP4_CAEEL
64	7	1.6	385	1 YFD0_YEAST
65	7	1.6	396	1 C121_MYCTU
66	7	1.6	401	1 NADM_YEAST
67	7	1.6	402	1 POXM_DROME
68	7	1.6	435	1 A2N1_HUMAN
69	7	1.6	438	1 SHIA_ECOLI
70	7	1.6	475	1 PERI_MOUSE
71	7	1.6	484	1 MEC2_MOUSE
72	7	1.6	486	1 MEC2_HUMAN
73	7	1.6	490	1 MOT3_YEAST
74	7	1.6	500	1 MOT1_HUMAN
75	7	1.6	505	1 MATK_SPIOL
76	7	1.6	513	1 GALT_BACSU
77	7	1.6	513	1 YAS1_METJA
78	7	1.6	513	1 YF65_METJA
79	7	1.6	515	1 PP2_SCHPO
80	7	1.6	554	1 DHAB_SALTY
81	7	1.6	611	1 SNF1_CANGA
82	7	1.6	619	1 SNF1_CANTR
83	7	1.6	623	1 PNT1_DROME
84	7	1.6	680	1 RPOC_TOBAC
85	7	1.6	682	1 SNK_MOUSE
86	7	1.6	682	1 SNK_RAT
87	7	1.6	685	1 SNK_HUMAN
88	7	1.6	753	1 YBV2_YEAST
89	7	1.6	778	1 TOP1_CANAL
90	7	1.6	825	1 IL4R_HUMAN
91	7	1.6	843	1 AXN2_HUMAN
92	7	1.6	956	1 PMA1_LYCES
93	7	1.6	956	1 PMA4_ARATH
94	7	1.6	960	1 PMA4_ARATH
95	7	1.6	1096	1 RPOB_GUTH
96	7	1.6	1115	1 E2K3_HUMAN
97	7	1.6	1178	1 PHYB_SORBI
98	7	1.6	1244	1 YP83_CAEEL
99	7	1.6	1342	1 Z335_HUMAN
100	7	1.6	1411	1 TCOF_HUMAN
101	7	1.6	1597	1 SOL_DROME
102	6	1.4	15	1 TAL_TREBR
103	6	1.4	16	1 LPH1_ECOLI
104	6	1.4	47	1 YQ54_BACAN
105	6	1.4	72	1 GBGE_CALVI
106	6	1.4	72	1 GBGE_DROME

Q9wt5	acomys	cahi
Q9wur3	akodon	jels
Q9dpx3	akodon	jels
Q9wux6	uranomys	ru
Q4251	helicobacte	
P20710	bacterioph	
P48298	mus musculu	
P97545	rattus norv	
Q0657	bubalus bub	
Q29542	giraffa cam	
Q58888	methanococ	
Q02383	homo sapien	
P93804	zea mays (m	
P93805	zea mays (m	
P36026	saccharomyc	
P30189	drosophila	
P51862	saccharomyc	
Q9hmv9	halobacteri	
Q92im7	staphylococ	
P44777	haemophilus	
Q64279	m heart- an	
P97832	rattus norv	
P09070	xenopus lae	
Q9cka5	pasteurella	
P39075	bacillus su	
Q10591	mycobacteri	
Q09461	caenorhabdi	
Q13750	schizosacch	
Q9183	thermotoga	
Q10948	caenorhabdi	
P43567	saccharomyc	
Q59571	mycobacteri	
Q06178	saccharomyc	
P23757	drosophila	
P20172	homo sapien	
P76350	escherichia	
P15331	mus musculu	
Q922d6	mus. musculu	
P51608	homo sapien	
P54785	saccharomyc	
P53985	homo sapien	
Q9m3n0	spinacia ol	
P35751	bacillus su	
Q58451	methanococ	
Q58960	methanococ	
P78968	schizosacch	
P37450	salmonella	
Q00372	candida gla	
Q94168	candida tro	
P51022	drosophila	
P12116	nicotiana t	
P53351	mus musculu	
Q9r012	rattus norv	
Q9nyv3	homo sapien	
P38261	saccharomyc	
Q00313	candida alb	
P24394	homo sapien	
Q9y2t1	homo sapien	
P22180	lycopersico	
Q91v11	arabidopsis	
Q9su58	arabidopsis	
Q78485	guillardia	
Q9n2j5	homo sapien	
P93527	sorghum bic	
Q09441	caenorhabdi	
Q9h422	homo sapien	
Q13428	homo sapien	
P27398	drosophila	
P34070	tremella br	
P3058	escherichia	
Q9rmx8	bacillus an	
Q9n1z2	calliphora	
Q9n1z3	drosophila	

107	6	1.4	76	1	CSRA_HELPJ	Q9zjh4	helicobacte	180	6	1.4	175	1	UL21_HCMVA	P16759	human cytom
108	6	1.4	76	1	CSRA_HELPJ	Q25983	helicobacte	181	6	1.4	176	1	FRIM_SALSA	P49947	salmo salar
109	6	1.4	77	1	TAT1_AQUAE	O66478	aquifex aeo	182	6	1.4	177	1	IPYR_NEIMA	Q9jvg3	neisseria m
110	6	1.4	83	1	CALD_BOVIN	Q27976	bos taurus	183	6	1.4	177	1	IPYR_NEIMB	Q9kq4	neisseria m
111	6	1.4	88	1	TH12_DICDI	P29446	dictyosteli	184	6	1.4	180	1	HYCF_ECOLI	P16432	escherichia
112	6	1.4	88	1	Y600_THEAC	Q9hkj8	thermoplasm	185	6	1.4	180	1	PYRE_METTH	O27888	methanobact
113	6	1.4	89	1	Y625_THEVO	Q97b34	thermoplasm	186	6	1.4	181	1	HYFH_ECOLI	P77423	escherichia
114	6	1.4	90	1	RPOL_SULAC	P46217	sulfolobus	187	6	1.4	182	1	YNH3_SULTO	Q96xy5	sulfolobus
115	6	1.4	91	1	GATC_BORBU	O51318	borrelia bu	188	6	1.4	183	1	ATPD_RICFR	Q9zcf2	rickettsia
116	6	1.4	94	1	EX7S_RALSO	O8xx93	ralstonia s	189	6	1.4	183	1	Y829_ARCFU	Q29429	archaeoglob
117	6	1.4	94	1	V093_FOWPV	Q9j5c7	fowlpox vir	190	6	1.4	185	1	SLVD_HELPJ	Q9zK89	helicobacte
118	6	1.4	95	1	GLR4_RICPR	Q9zdw1	rickettsia	191	6	1.4	185	1	SLVD_HELPJ	Q25748	helicobacte
119	6	1.4	96	1	VE4_HPV35	P27224	human papil	192	6	1.4	185	1	SOMA_KATPE	P20391	katsuwonus
120	6	1.4	102	1	ULC5_HCMVA	P16835	human cytom	193	6	1.4	188	1	Y101_UREPA	Q9pr43	ureaplasma
121	6	1.4	105	1	THIO_RICCN	Q921r5	rickettsia	194	6	1.4	189	1	YTRE_LEPBI	P20464	leptospira
122	6	1.4	105	1	THIO_RICPR	Q9zee0	rickettsia	195	6	1.4	190	1	HUNE_DROCR	O46236	drosophila
123	6	1.4	106	1	THIO_STRAU	P33791	streptomyce	196	6	1.4	190	1	YEIP_ECOLI	P33028	escherichia
124	6	1.4	107	1	LSM7_YEAST	P53905	saccharomyc	197	6	1.4	191	1	CLPI_CHLMU	Q9pjw1	chlamydia t
125	6	1.4	108	1	RBS_NITVU	Q59614	nitrobacter	198	6	1.4	191	1	CLPI_CHLTR	P38002	chlamydia t
126	6	1.4	110	1	YJ20_YEAST	P47092	saccharomyc	199	6	1.4	191	1	MOBA_PYRAB	Q9v0d0	pyrococcus
127	6	1.4	112	1	VE4_HPV29	P51897	human papil	200	6	1.4	192	1	HUNE_DROAD	O46232	drosophila
128	6	1.4	113	1	VG6_SPV1R	P15897	spiroplasma	201	6	1.4	192	1	HUNE_DROTA	O46360	drosophila
129	6	1.4	114	1	ASR2_LYCES	P37219	lycopersico	202	6	1.4	192	1	YT72_PSEAE	Q9hcn2	pseudomonas
130	6	1.4	116	1	Y472_METJA	Q58472	methanococc	203	6	1.4	193	1	HUNE_DROIK	O46242	drosophila
131	6	1.4	117	1	Y286_METJA	Q57734	methanococc	204	6	1.4	195	1	HUNE_DRODA	O46262	drosophila
132	6	1.4	120	1	SY16_HUMAN	O15467	h small ind	205	6	1.4	195	1	KITH_BACNA	O52951	bacillus su
133	6	1.4	121	1	GLB_TETPY	P17724	tetrahymena	206	6	1.4	196	1	HUNE_DROAA	O46234	drosophila
134	6	1.4	124	1	RL18_AQUAE	O67564	aquifex aeo	207	6	1.4	196	1	HUNE_DROSL	O46256	drosophila
135	6	1.4	124	1	RNPA_CAVPO	P00678	cavia porce	208	6	1.4	196	1	RR4_ALOPR	P36445	alopecurus
136	6	1.4	124	1	RNP_BALAC	P00673	balaeopter	209	6	1.4	196	1	SOMA_FUGRU	O12980	fugu rubrip
137	6	1.4	124	1	RNP_CHIBR	P00675	chinchilla	210	6	1.4	198	1	HUNE_DROCY	O46238	drosophila
138	6	1.4	124	1	RNP_GALMU	P00680	galea muste	211	6	1.4	198	1	HUNE_DRODS	O46240	drosophila
139	6	1.4	125	1	YO07_BPL2	P42542	bacterioph	212	6	1.4	198	1	HUNE_DROLI	O46244	drosophila
140	6	1.4	128	1	RNP_CHOHO	P00685	choleopus h	213	6	1.4	199	1	HIS5_HAEIN	P44340	haemophilus
141	6	1.4	128	1	RNP_PREEN	P19644	presbytis e	214	6	1.4	199	1	RS4_BACSU	P21466	bacillus su
142	6	1.4	136	1	GLB3_CHITP	P22431	chironomus	215	6	1.4	200	1	PHNE_BRUME	O8y221	brucella me
143	6	1.4	140	1	PRO3_WHEAT	P49234	triticum ae	216	6	1.4	200	1	PHNE_PSEAE	P09786	pseudomonas
144	6	1.4	141	1	PRO2_WHEAT	P49233	triticum ae	217	6	1.4	202	1	HANI_CHICK	Q08691	gallus gall
145	6	1.4	143	1	RNBR_SHEEP	Q29543	ovis aries	218	6	1.4	204	1	HANI_SHEEP	Q28555	ovis aries
146	6	1.4	144	1	Y970_METJA	O58380	methanococc	219	6	1.4	204	1	SOMA_SCIOC	Q91b11	sciaenops o
147	6	1.4	144	1	Y991_METJA	O58398	methanococc	220	6	1.4	208	1	ATPF_MYCGE	P47643	mycoplasma
148	6	1.4	144	1	YB8A_YEAST	P38354	saccharomyc	221	6	1.4	210	1	KTHV_SCHPO	P36590	schizosacch
149	6	1.4	146	1	PA21_PIG	P00592	sus scrofa	222	6	1.4	210	1	SOMI_ONCMY	P09538	oncorhynch
150	6	1.4	147	1	Y170_UREPA	Q9pxk3	ureaplasma	223	6	1.4	210	1	SOMI_ONCNE	Q91222	oncorhynch
151	6	1.4	149	1	Y919_METJA	O58329	methanococc	224	6	1.4	210	1	SOM2_ONCMY	P20332	oncorhynch
152	6	1.4	150	1	YM34_YEAST	Q03818	saccharomyc	225	6	1.4	210	1	SOM2_ONCNE	Q91221	oncorhynch
153	6	1.4	151	1	GLB3_CHITH	P02229	chironomus	226	6	1.4	210	1	SOMA_CORAU	P45655	coregonus a
154	6	1.4	151	1	RNBR_CAPCA	P79351	capreolus c	227	6	1.4	210	1	SOMA_ONCKE	P07064	oncorhynch
155	6	1.4	156	1	RNP_CAUCR	O9a594	caulobacter	228	6	1.4	210	1	SOMA_ONCKI	P10807	oncorhynch
156	6	1.4	156	1	RNP_HUMAN	P07998	homo sapien	229	6	1.4	210	1	SOMA_ONCMA	Q9d9g5	oncorhynch
157	6	1.4	156	1	RNP_MYOGL	Q9wus1	myoxus glis	230	6	1.4	210	1	SOMA_ONCTS	Q07221	oncorhynch
158	6	1.4	158	1	HUNE_DROMM	O46248	drosophila	231	6	1.4	210	1	SOMA_SALSA	P10814	salmo salar
159	6	1.4	158	1	ISPF_BACSU	Q06756	bacillus su	232	6	1.4	210	1	YRL1_METVA	P15886	methanococc
160	6	1.4	159	1	HUNE_DROSO	O46258	drosophila	233	6	1.4	212	1	PECF_MASLA	Q29730	mastigoclad
161	6	1.4	161	1	Y709_PYRHO	O58440	pyrococcus	234	6	1.4	213	1	RS6E_SULTO	Q97sn7	sulfolobus
162	6	1.4	161	1	Y882_PYRAB	Q9uz20	pyrococcus	235	6	1.4	214	1	RS6E_SULSO	Q980a6	sulfolobus
163	6	1.4	162	1	PHA3_FREDI	P14876	fremyella d	236	6	1.4	214	1	YJ82_AQUAE	O67790	aquifex aeo
164	6	1.4	163	1	ATPX_GUITH	Q78478	guillardia d	237	6	1.4	215	1	HANI_HUMAN	O96004	homo sapien
165	6	1.4	163	1	DCMS_HVDPS	P19915	hydrogenoph	238	6	1.4	215	1	HANI_RABIT	P57100	oryctolagus
166	6	1.4	164	1	TSTP_HSVSA	P18347	herpesvirus	239	6	1.4	215	1	PCP_STRPY	Q01328	streptococc
167	6	1.4	165	1	2A5E_RABIT	Q28654	o serine/th	240	6	1.4	215	1	UL14_HSV11	P04291	herpes simp
168	6	1.4	167	1	CHEW_ECOLI	P07365	escherichia	241	6	1.4	216	1	HIS5_AGR5	P58788	agrobacteri
169	6	1.4	167	1	CHEW_SALTY	P06110	salmonella	242	6	1.4	216	1	HIS5_RHIME	Q92tb1	rhizobium m
170	6	1.4	167	1	RNBR_BOVIN	P39873	bos taurus	243	6	1.4	216	1	YQFN_BACSU	P54471	bacillus su
171	6	1.4	168	1	DYR_STRPN	O54801	streptococc	244	6	1.4	217	1	YD67_SCHPO	O10319	schizosacch
172	6	1.4	169	1	GBP_XENLA	O93343	xenopus lae	245	6	1.4	218	1	BS25_BOVIN	Q27957	bos taurus
173	6	1.4	169	1	WBDQ_ECO57	O85341	escherichia	246	6	1.4	218	1	RNS7_NICAL	Q40381	nicotiana a
174	6	1.4	171	1	HUNE_SCAAL	O46254	scaptomyza	247	6	1.4	218	1	Y154_UREPA	Q9pqz0	ureaplasma
175	6	1.4	174	1	IPYR_THETH	P38576	thermus the	248	6	1.4	225	1	GS27_ARATH	Q9sj16	arabidopsis
176	6	1.4	174	1	NU6M_CERCA	Q34050	ceratitis c	249	6	1.4	226	1	CHEA_ENTAE	P21813	enterobacte
177	6	1.4	174	1	NU6M_DROME	P18933	drosophila	250	6	1.4	228	1	TSN_CRIGR	P78891	cricetulus
178	6	1.4	174	1	NU6M_DROYA	P07709	drosophila	251	6	1.4	228	1	TSN_HUMAN	Q15631	homo sapien
179	6	1.4	175	1	DEF_CAMJE	Q9pit8	campylobact	252	6	1.4	228	1	TSN_MOUSE	Q62348	mus musculu

253	1.4	229	1	CLPP_MESVI	Q9muu8 mesostigma	326	1.4	303	1	MOX2_HUMAN	P50222 homo sapien
254	1.4	229	1	SFSA_THETN	Q8r7n6 thermoanaer	327	1.4	303	1	MOX2_MOUSE	P32443 mus musculus
255	1.4	230	1	DHOV_HUMAN	P16083 homo sapien	328	1.4	303	1	MOX2_RAT	P39020 rattus norv
256	1.4	231	1	CEL_MAIZE	Q08704 zea mays (m	329	1.4	304	1	BCHL_RHOCA	P26237 rhodobacter
257	1.4	231	1	RUL_BUCAI	P57149 buchnera ap	330	1.4	304	1	GSH2_HUMAN	Q9bzm3 homo sapien
258	1.4	231	1	TRYP_PIG	P00761 sus scrofa	331	1.4	305	1	GSH2_MOUSE	P31316 mus musculus
259	1.4	233	1	FLPA_SULTO	Q971w2 sulfolobus	332	1.4	306	1	RBSK_HAEIN	P44331 haemophilus
260	1.4	235	1	CAVL_CAEEL	Q94051 caenorhabdi	333	1.4	306	1	SURL_MOUSE	P09925 mus musculus
261	1.4	237	1	TIRA_HUMAN	P58753 homo sapien	334	1.4	306	1	SURL_RAT	Q9qxu2 rattus norv
262	1.4	237	1	YC28_PORPU	P51342 porphyra pu	335	1.4	307	1	YK55_YEAST	P36155 saccharomyc
263	1.4	238	1	AROD_STAAM	Q95vh7 staphylococ	336	1.4	308	1	YR03_CAEEL	Q10049 caenorhabdi
264	1.4	239	1	CU58_HUMAN	P58505 homo sapien	337	1.4	310	1	ARCC_HAEIN	P44769 haemophilus
265	1.4	239	1	RP35_BACTK	P26763 bacillus th	338	1.4	310	1	LDB_BACHD	Q9k5z8 bacillus ha
266	1.4	239	1	RPSE_BACSU	P06222 bacillus su	339	1.4	312	1	DHKK_MOUSE	Q70503 mus musculus
267	1.4	240	1	Y162_HELPJ	Q9zmr9 helicobacte	340	1.4	315	1	OA2_HUMAN	O60403 homo sapien
268	1.4	241	1	Y162_HELPJ	O24970 helicobacte	341	1.4	316	1	MLP_SCVMI	P01546 saccharomyc
269	1.4	241	1	YF89_METJA	O58984 methanococc	342	1.4	316	1	SRG8_CAEEL	P45565 caenorhabdi
270	1.4	246	1	D5OC_THEAC	Q9hkb7 thermoplasm	343	1.4	316	1	Y054_MYCGE	P47300 mycoplasma
271	1.4	246	1	SRIA_PHYPO	P09350 physarum po	344	1.4	317	1	FLGL_ECOLI	P29744 escherichia
272	1.4	246	1	YDS3_SCHPO	O14179 schizosacch	345	1.4	317	1	YS11_BORBU	P70838 borrelia bu
273	1.4	247	1	FMT_VIBAL	O87726 vibrio algi	346	1.4	318	1	ANX4_BOVIN	P13214 bos taurus
274	1.4	247	1	TRY2_CANFA	P06872 canis famil	347	1.4	318	1	OAHL_HUMAN	Q9v4a9 homo sapien
275	1.4	247	1	YCKI_BACSU	P39456 bacillus su	348	1.4	320	1	YAGO_ECOLI	P77183 escherichia
276	1.4	248	1	YV28_SCHPO	O13872 schizosacch	349	1.4	320	1	CH36_CERCA	P17110 ceratitisc
277	1.4	253	1	PCRB_METJA	Q58647 methanococc	350	1.4	321	1	ABRA_PLAFF	P23746 plasmodium
278	1.4	256	1	BLAB_BACCE	P14488 bacillus ce	351	1.4	321	1	CRFB_XENLA	Q91653 xenopus lae
279	1.4	257	1	ATP6_COCHE	P14862 cochliobol	352	1.4	321	1	PYRB_MYCLE	Q9cgr5 mycobacteri
280	1.4	257	1	BLA2_BACCE	P04190 bacillus ce	353	1.4	321	1	YH28_YEAST	P38884 saccharomyc
281	1.4	257	1	BLA2_BACSP	P10425 bacillus sp	354	1.4	322	1	CRFB_HUMAN	P24387 homo sapien
282	1.4	260	1	ER29_RAT	P52555 rattus norv	355	1.4	322	1	CRFB_MOUSE	Q60571 mus musculus
283	1.4	261	1	DHG1_BACME	P39482 bacillus me	356	1.4	322	1	CRFB_RAT	P23488 rattus norv
284	1.4	261	1	DHG2_BACME	P39483 bacillus me	357	1.4	323	1	CO3_XENLA	P23667 xenopus lae
285	1.4	261	1	DHG3_BACME	P39484 bacillus me	358	1.4	323	1	MAFB_HUMAN	Q9v5q3 homo sapien
286	1.4	261	1	DHG4_BACME	P39485 bacillus me	359	1.4	323	1	MAFB_MOUSE	P54841 mus musculus
287	1.4	261	1	DHGA_BACME	P10528 bacillus me	360	1.4	323	1	MAFB_RAT	P54842 rattus norv
288	1.4	261	1	DHG_BACME	P40288 bacillus me	361	1.4	323	1	OTX1_BRARE	Q91994 brachydanio
289	1.4	261	1	DHG_BACSU	P12310 bacillus su	362	1.4	323	1	YJIK_ECOLI	P39382 escherichia
290	1.4	261	1	MODA_MYCTU	P95157 mycobacteri	363	1.4	324	1	CRFB_SHEEP	Q28557 ovis aries
291	1.4	261	1	MOTB_BACSU	P28612 bacillus su	364	1.4	324	1	Y089_BORBU	O51116 borrelia bu
292	1.4	262	1	DHGB_BACME	P07999 bacillus me	365	1.4	325	1	PLSX_MYCPU	Q98r48 mycoplasma
293	1.4	262	1	MURI_BUCAI	P57619 buchnera ap	366	1.4	326	1	PIX2_XENLA	Q9pwr3 xenopus lae
294	1.4	262	1	Y009_MYCGE	P47255 mycoplasma	367	1.4	326	1	SCFR_PEDPE	P3472 pediococcus
295	1.4	262	1	YLU4_CAEEL	P34398 caenorhabdi	368	1.4	326	1	YF87_CAEEL	Q03640 caenorhabdi
296	1.4	265	1	CMFF_SCHPO	P78794 schizosacch	369	1.4	327	1	ANX8_MOUSE	Q35640 mus musculus
297	1.4	269	1	ML22_LACLC	P50178 lactococcus	370	1.4	328	1	CYF_SYNY3	P26287 synecocyst
298	1.4	270	1	SFR5_MOUSE	Q53326 mus musculus	371	1.4	328	1	DLX2_HUMAN	Q07687 homo sapien
299	1.4	271	1	DB83_CAEEL	Q9xwv0 caenorhabdi	372	1.4	329	1	SRP1_YEAST	P38271 saccharomyc
300	1.4	271	1	HW8_XENLA	P14837 xenopus lae	373	1.4	329	1	SRPG_SYNP7	O59966 synecococc
301	1.4	272	1	HXA9_HUMAN	P31269 homo sapien	374	1.4	330	1	PGI1_ARATH	Q9m5j9 arabidopsis
302	1.4	273	1	CXB5_HUMAN	O95377 homo sapien	375	1.4	330	1	VP19_HSVSA	Q01051 herpesvirus
303	1.4	273	1	DHOV_MOUSE	Q64669 mus musculus	376	1.4	331	1	HXA1_MOUSE	P09022 mus musculus
304	1.4	274	1	GADX_ECO27	Q9eyv5 escherichia	377	1.4	331	1	LDHA_HUMAN	P00338 homo sapien
305	1.4	274	1	GADX_ECO57	P58230 escherichia	378	1.4	331	1	LDHA_MACFA	Q9be24 macaca fasc
306	1.4	274	1	GADX_ECOLI	P37639 escherichia	379	1.4	331	1	LDHA_PELSU	Q98s10 pelodiscus
307	1.4	274	1	YEO3_MYCTU	P71671 mycobacteri	380	1.4	331	1	LDHA_TRASC	Q9p4c3 trachemys s
308	1.4	276	1	YV01_METJA	Q60300 methanococc	381	1.4	332	1	DLX2_MOUSE	P40764 mus musculus
309	1.4	277	1	RIPA_LUFCE	Q00465 luffa cylin	382	1.4	332	1	FLIM_BACSU	P23453 bacillus su
310	1.4	280	1	SUMT_PSEDE	P21631 pseudomonas	383	1.4	332	1	LDHA_BRARE	Q9pvk5 brachydanio
311	1.4	280	1	YB80_YEAST	P32788 saccharomyc	384	1.4	333	1	HXA1_RAT	O08656 rattus norv
312	1.4	281	1	R33_MYCLE	Q32987 mycobacteri	385	1.4	334	1	REGA_CIOAB	P58258 clostridium
313	1.4	284	1	Y46B_MYCPN	Q50316 mycoplasma	386	1.4	335	1	DHYS_PYRAB	Q9v0n5 pyrococcus
314	1.4	285	1	Y048_HAEIN	P44481 haemophilus	387	1.4	335	1	HXA1_HUMAN	P49639 homo sapien
315	1.4	287	1	UPK3_MOUSE	Q91kx8 mus musculus	388	1.4	338	1	ANX2_BOVIN	P04272 bos taurus
316	1.4	288	1	VA08_VACCC	P20986 vaccinia vi	389	1.4	338	1	ANX2_CHICK	P17785 gallus gall
317	1.4	288	1	VA08_VARV	P33834 variola vir	390	1.4	338	1	ANX2_MOUSE	P07356 mus musculus
318	1.4	289	1	APAH_YERPE	Q8z1k7 versinia pe	391	1.4	338	1	SD22_YEAST	P36047 saccharomyc
319	1.4	293	1	ERA_MYCPU	Q98qil mycoplasma	392	1.4	339	1	SYR2_RHIN	P57333 rhizobium s
320	1.4	293	1	URER_PROMI	Q02458 proteus mir	393	1.4	339	1	YU99_YEAST	P47163 saccharomyc
321	1.4	295	1	MCBB_ECOLI	P23184 escherichia	394	1.4	341	1	G3P2_CAEEL	P17329 caenorhabdi
322	1.4	296	1	ARG1_BACSU	P39138 bacillus su	395	1.4	341	1	G3P3_CAEEL	P17330 caenorhabdi
323	1.4	298	1	MOX2_XENLA	P39021 xenopus lae	396	1.4	343	1	BIOR_BUCAI	P57378 buchnera ap
324	1.4	298	1	Y812_ARCFU	O29446 archaeoglob	397	1.4	344	1	HYE1_HYONI	P24397 hyoscyamus
325	1.4	302	1	HYPB_BRAJA	Q45257 bradyrhizob	398	1.4	346	1	HYE1_YEAST	P41834 saccharomyc

399	6	1.4	1	1	1	472	6	1.4	391	1	CAJ1_YEAST	P39101	saccharomyc
400	6	1.4	1	1	1	473	6	1.4	391	1	HERP_HUMAN	P22140	saccharomyc
401	6	1.4	1	1	1	474	6	1.4	391	1	PEP3_PEA	Q15011	homo sapien
402	6	1.4	1	1	1	475	6	1.4	391	1	PEP3_PEA	Q43064	pisum sativ
403	6	1.4	1	1	1	476	6	1.4	392	1	CARS_RHINI	P43232	rhizopus ni
404	6	1.4	1	1	1	477	6	1.4	392	1	ODD_DRONE	P23803	drosophila
405	6	1.4	1	1	1	478	6	1.4	393	1	THL_BACSU	P45855	bacillus su
406	6	1.4	1	1	1	479	6	1.4	394	1	VORA_PVRFU	Q57801	pyrococcus
407	6	1.4	1	1	1	480	6	1.4	398	1	AAAD_HUMAN	P22760	homo sapien
408	6	1.4	1	1	1	481	6	1.4	398	1	CATD_CHICK	Q05744	gallus gall
409	6	1.4	1	1	1	482	6	1.4	400	1	CINA_ECOLI	P77808	escherichia
410	6	1.4	1	1	1	483	6	1.4	400	1	YD01_METJA	Q58697	methanococc
411	6	1.4	1	1	1	484	6	1.4	403	1	DDC_DROLE	Q96571	drosophila
412	6	1.4	1	1	1	485	6	1.4	403	1	MAF_HUMAN	Q75444	homo sapien
413	6	1.4	1	1	1	486	6	1.4	403	1	TGT_ARCFU	Q28787	archaeoglob
414	6	1.4	1	1	1	487	6	1.4	404	1	SNX5_MOUSE	Q9Y5X3	homo sapien
415	6	1.4	1	1	1	488	6	1.4	404	1	GLYA_MYCPN	Q9D8U8	mus musculu
416	6	1.4	1	1	1	489	6	1.4	406	1	NUCM_DICDI	P78011	mycoplasma
417	6	1.4	1	1	1	490	6	1.4	406	1	APG3_METTH	Q23883	dictyosteli
418	6	1.4	1	1	1	491	6	1.4	409	1	CAQC_RABIT	Q27628	methanobact
419	6	1.4	1	1	1	492	6	1.4	409	1	Y291_METJA	P12355	oryctolagus
420	6	1.4	1	1	1	493	6	1.4	409	1	BR3B_HUMAN	Q57739	methanococc
421	6	1.4	1	1	1	494	6	1.4	410	1	CAQC_CANFA	Q12837	homo sapien
422	6	1.4	1	1	1	495	6	1.4	410	1	BR3B_MOUSE	P12637	canis famli
423	6	1.4	1	1	1	496	6	1.4	411	1	RAPS_HUMAN	Q63934	mus musculu
424	6	1.4	1	1	1	497	6	1.4	411	1	DJA2_HUMAN	Q13702	homo sapien
425	6	1.4	1	1	1	498	6	1.4	412	1	DJA2_MOUSE	Q60884	homo sapien
426	6	1.4	1	1	1	499	6	1.4	412	1	DJA2_MOUSE	Q9QV10	mus musculu
427	6	1.4	1	1	1	500	6	1.4	412	1	DJA2_MOUSE	Q5824	rattus norv
428	6	1.4	1	1	1	501	6	1.4	412	1	I5P1_CANFA	Q29467	canis famli
429	6	1.4	1	1	1	502	6	1.4	412	1	I5P1_HUMAN	Q14642	homo sapien
430	6	1.4	1	1	1	503	6	1.4	413	1	CAQC_RAT	P51868	rattus norv
431	6	1.4	1	1	1	504	6	1.4	413	1	CP13_RAT	P05544	rattus norv
432	6	1.4	1	1	1	505	6	1.4	413	1	FENR_SYNY3	Q55318	synthocyst
433	6	1.4	1	1	1	506	6	1.4	414	1	TYV1_MOUSE	P25490	homo sapien
434	6	1.4	1	1	1	507	6	1.4	414	1	TYV1_MOUSE	Q00899	mus musculu
435	6	1.4	1	1	1	508	6	1.4	415	1	CAQC_MOUSE	Q09161	mus musculu
436	6	1.4	1	1	1	509	6	1.4	415	1	REF1_METAC	Q8T900	methanosarc
437	6	1.4	1	1	1	510	6	1.4	415	1	RUN3_HUMAN	Q13761	h runt-rela
438	6	1.4	1	1	1	511	6	1.4	415	1	YW92_MYCTU	P96897	mycobacteri
439	6	1.4	1	1	1	512	6	1.4	415	1	Y207_METJA	Q60269	methanococc
440	6	1.4	1	1	1	513	6	1.4	418	1	CP16_RAT	P09006	rattus norv
441	6	1.4	1	1	1	514	6	1.4	421	1	BR3A_MOUSE	P17208	mus musculu
442	6	1.4	1	1	1	515	6	1.4	421	1	DM3L_MOUSE	Q9CWR8	mus musculu
443	6	1.4	1	1	1	516	6	1.4	421	1	SYS_MYCPN	Q98RH5	mycoplasma
444	6	1.4	1	1	1	517	6	1.4	422	1	CSD_BORBU	O51111	borrelia bu
445	6	1.4	1	1	1	518	6	1.4	422	1	OMD_BOVIN	O77742	bos taurus
446	6	1.4	1	1	1	519	6	1.4	423	1	AACT_HUMAN	P01011	homo sapien
447	6	1.4	1	1	1	520	6	1.4	423	1	BR3A_HUMAN	Q01851	homo sapien
448	6	1.4	1	1	1	521	6	1.4	423	1	PNT1_YEAST	P38969	saccharomyc
449	6	1.4	1	1	1	522	6	1.4	423	1	YJ54_YEAST	P47130	saccharomyc
450	6	1.4	1	1	1	523	6	1.4	425	1	L756_CAEEL	Q11184	caenorhabdi
451	6	1.4	1	1	1	524	6	1.4	425	1	POU1_BRARE	P31366	brachydanio
452	6	1.4	1	1	1	525	6	1.4	425	1	YF31_METJA	Q58926	methanococc
453	6	1.4	1	1	1	526	6	1.4	425	1	YNN6_YEAST	P53911	saccharomyc
454	6	1.4	1	1	1	527	6	1.4	428	1	FXB2_MOUSE	Q64733	mus musculu
455	6	1.4	1	1	1	528	6	1.4	428	1	MTBA_BACAR	P19888	bacillus an
456	6	1.4	1	1	1	529	6	1.4	429	1	K1CR_HUMAN	P05783	homo sapien
457	6	1.4	1	1	1	530	6	1.4	430	1	NOR2_RAT	Q63516	rattus norv
458	6	1.4	1	1	1	531	6	1.4	432	1	GBAL_CRYNE	P54853	cryptococcu
459	6	1.4	1	1	1	532	6	1.4	432	1	GLPB_HAEIN	P43800	haemophilus
460	6	1.4	1	1	1	533	6	1.4	434	1	YNN6_YEAST	P53916	saccharomyc
461	6	1.4	1	1	1	534	6	1.4	435	1	MALC_STRPN	Q04698	streptococc
462	6	1.4	1	1	1	535	6	1.4	435	1	XYLA_TETHA	O82845	tetragenoco
463	6	1.4	1	1	1	536	6	1.4	436	1	BRNQ_HAEIN	P71345	haemophilus
464	6	1.4	1	1	1	537	6	1.4	437	1	PR54_YEAST	P40327	saccharomyc
465	6	1.4	1	1	1	538	6	1.4	437	1	ZP12_BRARE	P56224	brachydanio
466	6	1.4	1	1	1	539	6	1.4	438	1	FUCP_ECOLI	P11551	escherichia
467	6	1.4	1	1	1	540	6	1.4	439	1	PR54_DROME	P48601	drosophila
468	6	1.4	1	1	1	541	6	1.4	439	1	SLAP_LACHE	P38059	lactobacilli
469	6	1.4	1	1	1	542	6	1.4	440	1	PR54_CHICK	Q90732	gallus gall
470	6	1.4	1	1	1	543	6	1.4	440	1	PR54_HUMAN	Q03527	homo sapien
471	6	1.4	1	1	1	544	6	1.4	441	1	BR11_BRARE	Q90436	brachydanio

545	6	1.4	441	1	FLID_AQUAE	O67805	equifex	ao	618	6	1.4	496	1	CHK1_SCHPO	P34208	schizosacch
546	6	1.4	441	1	GAT6_RAT	P46153	rattus	norv	619	6	1.4	496	1	NUSA_BUCAI	P57459	buchnera ap
547	6	1.4	442	1	VATC_DROME	Q9v7n5	drosophila		620	6	1.4	497	1	BRNL_RAT	O32262	rattus norv
548	6	1.4	443	1	DCUA_HELPY	O25425	helicobacte		621	6	1.4	498	1	SYTA_RAT	O08625	rattus norv
549	6	1.4	443	1	2P23_BRARE	P79745	brachydanio		622	6	1.4	500	1	BRNL_HUMAN	P20264	homo sapien
550	6	1.4	444	1	FXF2_HUMAN	Q12947	homo sapien		623	6	1.4	501	1	YSV1_CAEEL	Q22556	caenorhabdl
551	6	1.4	444	1	GAT6_MOUSE	K11169	mus musculus		624	6	1.4	504	1	VPS1_AHSV6	Q71026	african hor
552	6	1.4	447	1	AMPA_MYCGE	P47631	mycoplasma		625	6	1.4	505	1	CXAA_MOUSE	Q9wu54	mus musculus
553	6	1.4	448	1	OC26_HUMAN	Q03052	homo sapien		626	6	1.4	505	1	VPS1_AHSV9	Q96597	african hor
554	6	1.4	449	1	GAT6_HUMAN	Q92908	homo sapien		627	6	1.4	505	1	TDT_AMBME	O57486	ambystoma m
555	6	1.4	449	1	OC26_MOUSE	P21952	mus musculus		628	6	1.4	508	1	CROC_DROME	P32027	drosophila
556	6	1.4	450	1	DNAA_MYCCA	P24116	mycoplasma		629	6	1.4	508	1	SYM_RICPR	Q9z6P0	rickettsia
557	6	1.4	451	1	FXGB_CHICK	Q90964	gallus gall		630	6	1.4	510	1	FKH_DROME	P14734	drosophila
558	6	1.4	451	1	OC26_RAT	P20267	rattus norv		631	6	1.4	511	1	CP4B_RAT	P15129	rattus norv
559	6	1.4	452	1	DHAP_RAT	P11883	rattus norv		632	6	1.4	511	1	VGLC_HSV11	P10228	herpes simp
560	6	1.4	453	1	DHAP_HUMAN	P30838	homo sapien		633	6	1.4	511	1	VGLC_HSV1K	P28986	herpes simp
561	6	1.4	453	1	DHAP_MOUSE	P47739	mus musculus		634	6	1.4	512	1	RCK1_YEAST	P38622	saccharomyc
562	6	1.4	453	1	PLM2_PLAFA	P46925	plasmodium		635	6	1.4	513	1	AMY3_SCHPO	O14154	schizosacch
563	6	1.4	457	1	PRTC_STGRG	P23230	streptomyce		636	6	1.4	515	1	ALAB_MESAU	P18841	mesocricetu
564	6	1.4	461	1	NIFN_KLEPN	P08738	klebsiella		637	6	1.4	515	1	CRBA_DROME	P29747	drosophila
565	6	1.4	462	1	DRTS_PARTE	P27828	paramecium		638	6	1.4	517	1	CPN1_RANCA	Q92104	rana catesb
566	6	1.4	464	1	VNSS_TSW1	P26002	tomato spot		639	6	1.4	518	1	GPI1_YEAST	P25346	saccharomyc
567	6	1.4	465	1	HNF6_HUMAN	Q9ubc0	homo sapien		640	6	1.4	522	1	HMCS_CHICK	P23228	gallus gall
568	6	1.4	465	1	HNF6_MOUSE	O08755	mus musculus		641	6	1.4	523	1	RPN3_YEAST	P40016	saccharomyc
569	6	1.4	465	1	HNF6_RAT	P70512	rattus norv		642	6	1.4	523	1	SYTA_MOUSE	Q9r0M4	mus musculus
570	6	1.4	466	1	CF1A_DROME	P16241	drosophila		643	6	1.4	525	1	VE59_LAMBD	P03754	bacterioph
571	6	1.4	466	1	COQ6_CAEEL	O01884	caenorhabdi		644	6	1.4	529	1	RAI2_MOUSE	Q9qv78	mus musculus
572	6	1.4	466	1	2IC3_MOUSE	Q62521	mus musculus		645	6	1.4	529	1	YEJF_ECOLI	P33916	escherichia
573	6	1.4	467	1	2A5E_HUMAN	Q16537	h. serine/th		646	6	1.4	530	1	MATP_HUMAN	Q9umv9	homo sapien
574	6	1.4	467	1	SYE_BUCAI	P57173	buchnera ap		647	6	1.4	530	1	ZIC2_MOUSE	Q82500	mus musculus
575	6	1.4	467	1	UBPC_TSHPW	P26033	schizosacch		648	6	1.4	532	1	INV1_YEAST	P10594	saccharomyc
576	6	1.4	467	1	VNSS_SCHPO	Q9m647	arabidopsis		649	6	1.4	532	1	INV2_YEAST	P00724	saccharomyc
577	6	1.4	467	1	ZIC3_HUMAN	O60481	homo sapien		650	6	1.4	532	1	INV4_YEAST	P10596	saccharomyc
578	6	1.4	468	1	ARLY_ZYMO	Q9z660	zymomonas m		651	6	1.4	532	1	PUR9_HAEIN	P43852	h bifunctio
579	6	1.4	468	1	PERI_RAT	P21807	rattus norv		652	6	1.4	532	1	YRR5_CAEEL	Q09346	caenorhabdi
580	6	1.4	469	1	A2AC_DIDMA	P35405	didelphis m		653	6	1.4	532	1	ZIC2_HUMAN	Q95409	homo sapien
581	6	1.4	469	1	FXGA_HUMAN	P55315	homo sapien		654	6	1.4	533	1	CWFO_SCHPO	Q9p6r8	schizosacch
582	6	1.4	469	1	IARI_ARATH	Q9m647	arabidopsis		655	6	1.4	533	1	GTR2_CHICK	Q90592	gallus gall
583	6	1.4	470	1	LDL1_BACST	P11959	bacillus st		656	6	1.4	533	1	PUR9_PASMU	P57828	p bifunctio
584	6	1.4	471	1	SYC_SULTO	Q96yc3	sulfolobus		657	6	1.4	536	1	IMA5_HUMAN	O15131	homo sapien
585	6	1.4	473	1	KNOB_PLAFA	P13817	plasmodium		658	6	1.4	537	1	KLC1_MOUSE	O88447	mus musculus
586	6	1.4	474	1	ANA_DROME	Q26307	drosophila		659	6	1.4	537	1	SY63_DISOM	P24507	discopyge o
587	6	1.4	474	1	HOSNL_PENCH	O94225	penicillium		660	6	1.4	538	1	VGLF_MUMPM	P19716	mumps virus
588	6	1.4	477	1	FXGB_HUMAN	P53315	homo sapien		661	6	1.4	538	1	VGLF_MUMPM	P11236	mumps virus
589	6	1.4	477	1	PEN3_ADECC	Q65950	canine aden		662	6	1.4	538	1	VGLF_MUMPR	P09458	mumps virus
590	6	1.4	477	1	RMUC_YERPE	Q8za18	versinia pe		663	6	1.4	538	1	VGLF_MUMPS	P33481	mumps virus
591	6	1.4	477	1	RT3D_ACTPL	O08633	actinobacil		664	6	1.4	542	1	TCPE_CAEEL	P47209	caenorhabdi
592	6	1.4	480	1	FXGB_RAT	Q00939	rattus norv		665	6	1.4	545	1	VNCS_DSNDV	O71153	diatraea sa
593	6	1.4	480	1	TGN2_HUMAN	O43493	homo sapien		666	6	1.4	545	1	VNCS_JCDNV	Q90054	junonia coe
594	6	1.4	481	1	FXGB_MOUSE	Q09087	mus musculus		667	6	1.4	548	1	PHR1_CANAL	P43076	candida alb
595	6	1.4	482	1	CABL_METJA	Q58773	methanococc		668	6	1.4	548	1	URL1_HUMAN	Q9nwz5	homo sapien
596	6	1.4	482	1	NURM_CAEEL	Q93873	caenorhabdi		669	6	1.4	548	1	URL1_MOUSE	O71Y13	mus musculus
597	6	1.4	482	1	PUR8_SCHPO	O60105	schizosacch		670	6	1.4	549	1	EPD2_CANNA	O74137	candida mal
598	6	1.4	483	1	GATB_RICCN	Q92j76	rickettsia		671	6	1.4	551	1	CBX4_MOUSE	O55187	mus musculus
599	6	1.4	483	1	GATB_RICPR	Q93rell	rickettsia		672	6	1.4	556	1	KLC1_RAT	P37285	rattus norv
600	6	1.4	484	1	MURC_BUCAI	P57310	buchnera ap		673	6	1.4	556	1	PMGI_RICCO	P35493	ricinus com
601	6	1.4	485	1	GLC2_SOYBN	P04405	glycine max		674	6	1.4	556	1	Y281_MYCGE	P47523	mycoplasma
602	6	1.4	485	1	ONC2_HUMAN	O95948	homo sapien		675	6	1.4	557	1	Y369_MYCGE	P47609	mycoplasma
603	6	1.4	485	1	RNF8_HUMAN	O76064	homo sapien		676	6	1.4	558	1	CBX4_HUMAN	O00257	homo sapien
604	6	1.4	486	1	COXW_YEAST	P40086	saccharomyc		677	6	1.4	558	1	PAP_CANAL	O42617	candida alb
605	6	1.4	487	1	MALQ_CLOBU	Q92666	clostridium		678	6	1.4	558	1	Y322_MYCGE	P47564	mycoplasma
606	6	1.4	488	1	ENV_HTLIN	Q03816	human t-cel		679	6	1.4	561	1	RK_BOVIN	P28327	bos taurus
607	6	1.4	488	1	PMGI_PRUDU	O24246	prunus dulc		680	6	1.4	563	1	RK_HUMAN	Q15835	homo sapien
608	6	1.4	488	1	Y647_MYCTU	P69536	mycobacteri		681	6	1.4	563	1	Y288_CHLTR	O84290	chlamydia t
609	6	1.4	490	1	CHEL_HUMAN	O00409	homo sapien		682	6	1.4	564	1	RK_RAT	Q53651	rattus norv
610	6	1.4	492	1	FTSA_HELPY	O25629	helicobacte		683	6	1.4	566	1	STR_METJA	Q57689	methanococc
611	6	1.4	492	1	MEC2_RAT	Q00566	rattus norv		684	6	1.4	568	1	DISC_DROME	P23752	drosophila
612	6	1.4	493	1	VA18_VACCC	P20534	vaccinia vi		685	6	1.4	568	1	RS1_RICPR	Q9zd28	rickettsia
613	6	1.4	493	1	VA18_VACCV	P16712	vaccinia vi		686	6	1.4	569	1	AMY_STRVL	P22998	streptomyce
614	6	1.4	493	1	VA18_VARV	P33827	variola vir		687	6	1.4	569	1	KLC1_HUMAN	Q07866	homo sapien
615	6	1.4	494	1	ICSB_SHIFL	P33546	shigella fl		688	6	1.4	570	1	YEDQ_SALTY	Q8znc5	salmonella
616	6	1.4	495	1	BRN1_MOUSE	P31361	mus musculus		689	6	1.4	572	1	DPOL_BP003	Q37882	bacterioph
617	6	1.4	496	1	BAF1_KLUNA	P33293	kluyveromyc		690	6	1.4	572	1	DPOL_BP02	P19894	bacterioph

691	1	1.4	575	1	FLA2_CAMJE	P22251 campylobact	764	1	1.4	678	1	SIRI_YEAST	P21691 saccharomyc
692	6	1.4	575	1	ILVB_LACLA	Q02137 lactococcus	765	1	1.4	680	1	LAG2_YEAST	Q92325 saccharomyc
693	6	1.4	576	1	DEAF_DROME	Q24180 drosophila	766	6	1.4	684	1	EP84_HCNVA	P17151 human cytom
694	6	1.4	578	1	AC22_STRCO	P46105 streptomyce	767	6	1.4	685	1	YG04_YEAST	P53118 saccharomyc
695	6	1.4	579	1	G160_HUMAN	Q08378 homo sapien	768	6	1.4	686	1	KLC_STRPU	Q05090 strongyloce
696	6	1.4	580	1	ATPU_YEAST	P22136 saccharomyc	769	6	1.4	687	1	HDAl_SCHPO	P56523 schizosacch
697	6	1.4	585	1	YHD9_YEAST	P38732 saccharomyc	770	6	1.4	688	1	EPG_MYCCE	P47335 mycoplasma
698	6	1.4	586	1	VATA_HALVO	Q48332 halobacteri	771	6	1.4	690	1	VGLH_HSV7J	P52353 human herpe
699	6	1.4	586	1	YS76_YEAST	P46982 saccharomyc	772	6	1.4	695	1	EPGL_SYNY3	P28371 synechocyst
700	6	1.4	587	1	SYT3_MOUSE	Q35681 mus musculu	773	6	1.4	696	1	YI95_SYNY3	P74101 synechocyst
701	6	1.4	588	1	SYT3_RAT	P40748 rattus norv	774	6	1.4	699	1	HS82_ORYSA	P33126 oryza sativ
702	6	1.4	590	1	PRIM_CHLPN	Q926w4 chlamydia p	775	6	1.4	705	1	CNG3_BOVIN	Q29441 bos taurus
703	6	1.4	590	1	SYT3_HUMAN	Q9b9g1 homo sapien	776	6	1.4	706	1	SEM2_MACMU	Q95196 macaca mula
704	6	1.4	594	1	DNBK_LYCES	O04059 lycopersico	777	6	1.4	707	1	ORC1_SCHPO	P47899 schizosacch
705	6	1.4	595	1	DNBK_MYCCE	P47547 mycoplasma	778	6	1.4	708	1	GIT2_MOUSE	Q91402 mus musculu
706	6	1.4	595	1	DNBK_MYCPN	P75344 mycoplasma	779	6	1.4	708	1	VP40_HCNVA	P16753 human cytom
707	6	1.4	596	1	HNEN_ANOGA	O02491 anopheles g	780	6	1.4	711	1	CAD2_LISMO	Q60048 listeria mo
708	6	1.4	597	1	STAR_DROME	P42519 drosophila	781	6	1.4	715	1	HEL5_PYRHO	O59025 pyrococcus
709	6	1.4	597	1	SYK_CRILLO	P37879 cricetulus	782	6	1.4	717	1	EMLI_HUMAN	Q00423 homo sapien
710	6	1.4	598	1	NR42_HUMAN	P43354 homo sapien	783	6	1.4	717	1	PRDD_HUMAN	Q9h4q3 homo sapien
711	6	1.4	598	1	NR42_MOUSE	Q06219 mus musculu	784	6	1.4	726	1	RNR_MYCPN	P75529 mycoplasma
712	6	1.4	598	1	NR42_RAT	Q07917 rattus norv	785	6	1.4	728	1	EF2_ARCFU	Q28385 archaeoglob
713	6	1.4	600	1	DHSA_PARDE	Q59661 paracoccus	786	6	1.4	728	1	MYBA_XENLA	Q05935 xenopus lae
714	6	1.4	600	1	Y018_MYCCE	P47264 mycoplasma	787	6	1.4	730	1	DCOR_LACS3	P43099 lactobacill
715	6	1.4	601	1	BS4_HUMAN	Q9y357 homo sapien	788	6	1.4	730	1	EF2_METTE	Q93640 methanosarc
716	6	1.4	601	1	PDM1_DROME	P31368 drosophila	789	6	1.4	732	1	EF2_PYRAB	Q9v128 pyrococcus
717	6	1.4	602	1	CRK_DAUCA	P53681 daucus caro	790	6	1.4	732	1	EF2_PYRFU	P29050 pyrococcus
718	6	1.4	603	1	MTM1_MOUSE	Q922C5 mus musculu	791	6	1.4	732	1	EF2_PYRHO	O59521 pyrococcus
719	6	1.4	604	1	DNBK_BACME	P05646 bacillus me	792	6	1.4	733	1	PURL_METJA	O58660 methanococ
720	6	1.4	604	1	XJ13_YEAST	P47030 saccharomyc	793	6	1.4	738	1	V047_HSV62	P52549 human herpe
721	6	1.4	606	1	DNBK_BACTR	Q9kws7 bacillus th	794	6	1.4	740	1	GAG_SMRVH	P21411 squirrel mo
722	6	1.4	606	1	HNMD_DROAN	P22544 drosophila	795	6	1.4	741	1	YD35_MYCPN	P75443 mycoplasma
723	6	1.4	608	1	EX5A_ECOLI	P04993 escherichia	796	6	1.4	743	1	ABRA_PLAFC	P22620 plasmodium
724	6	1.4	609	1	OPA_DROME	P39768 drosophila	797	6	1.4	745	1	Y0A4_CAEEL	Q10040 caenorhabdi
725	6	1.4	610	1	DNBK_BACSH	O69268 bacillus sp	798	6	1.4	748	1	MUTA_MOUSE	P16332 mus musculu
726	6	1.4	610	1	GIDA_CHLTR	O84506 chlamydia t	799	6	1.4	755	1	P100_HSV7J	P52519 human herpe
727	6	1.4	611	1	ADAS_DICDI	Q96759 dictyostell	800	6	1.4	757	1	ECR_LUCCU	O18531 lucilia cup
728	6	1.4	611	1	GIDA_CHLPN	Q9z7t7 chlamydia p	801	6	1.4	757	1	HUNB_DROSE	O62538 drosophila
729	6	1.4	618	1	DNBK_CYACA	Q9tlt1 cyanidium c	802	6	1.4	758	1	HUNB_DROME	P05084 drosophila
730	6	1.4	618	1	YKR4_YEAST	P36029 saccharomyc	803	6	1.4	758	1	METE_BUCAI	P57142 buchera ap
731	6	1.4	620	1	SNF1_CANAL	P52497 candida alb	804	6	1.4	758	1	SC18_YEAST	P18759 saccharomyc
732	6	1.4	621	1	CRT1_CERNC	P48537 cercospora	805	6	1.4	759	1	GIT2_HUMAN	Q14161 homo sapien
733	6	1.4	622	1	MTL1_LACLA	P35516 lactococcus	806	6	1.4	759	1	HUNB_DROYA	O62541 drosophila
734	6	1.4	626	1	NR43_HUMAN	Q92570 homo sapien	807	6	1.4	760	1	MCMT_SCHPO	O75001 schizosacch
735	6	1.4	628	1	NR43_RAT	P51179 rattus norv	808	6	1.4	763	1	DYRA_HUMAN	Q13627 homo sapien
736	6	1.4	631	1	DXS_STRC1	Q9rbn6 streptomyce	809	6	1.4	763	1	DYRA_MOUSE	Q61214 mus musculu
737	6	1.4	631	1	RPSD_BORBU	P52323 borrelia bu	810	6	1.4	763	1	DYRA_RAT	Q63470 rattus norv
738	6	1.4	632	1	ASNB_BACSU	P54420 bacillus su	811	6	1.4	763	1	IF2C_PORPU	P51257 porphyra pu
739	6	1.4	633	1	SNFL_YEAST	Q06782 saccharomyc	812	6	1.4	764	1	PAG_BACAN	P13423 bacillus an
740	6	1.4	634	1	KNOB_PLAFC	P09346 plasmodium	813	6	1.4	767	1	AMYH_SACDI	P04065 saccharomyc
741	6	1.4	634	1	MET4_YEAST	P32389 saccharomyc	814	6	1.4	767	1	CUL1_SCHPO	O13790 schizosacch
742	6	1.4	634	1	SELB_MOOTH	Q46455 moorella th	815	6	1.4	767	1	HUNB_DROOR	O62537 drosophila
743	6	1.4	636	1	NAF1_HUMAN	Q15025 homo sapien	816	6	1.4	768	1	AMY1_SACDI	P29760 saccharomyc
744	6	1.4	640	1	GYRB_SPICI	P34031 spiroplasma	817	6	1.4	770	1	TOPI1_THEAC	Q9hm08 thermoplas
745	6	1.4	642	1	ARE2_YEAST	P53629 saccharomyc	818	6	1.4	771	1	DNK3_SYNY3	P73098 synechocyst
746	6	1.4	643	1	HS71_DROME	P02825 drosophila	819	6	1.4	771	1	YKY0_YEAST	Q02208 saccharomyc
747	6	1.4	643	1	RO60_CAEEL	Q27274 caenorhabdi	820	6	1.4	772	1	PMIP_YEAST	P35999 saccharomyc
748	6	1.4	643	1	VP4B_VACCV	P06440 vaccinia vi	821	6	1.4	775	1	ANR5_MOUSE	Q9d217 mus musculu
749	6	1.4	644	1	VP4B_VACCC	P20643 vaccinia vi	822	6	1.4	776	1	ANR5_HUMAN	O9nu02 homo sapien
750	6	1.4	644	1	VP4B_VARV	P33818 variola vir	823	6	1.4	785	1	YG51_YEAST	P50089 saccharomyc
751	6	1.4	644	1	YEAG_ECOLI	P77351 escherichia	824	6	1.4	790	1	AD30_HUMAN	Q9ukf2 homo sapien
752	6	1.4	648	1	Y084_HUMAN	Q14699 homo sapien	825	6	1.4	791	1	LON_CAMJE	O69300 campylobact
753	6	1.4	653	1	YTH3_CAEEL	P54002 caenorhabdi	826	6	1.4	802	1	MUTS_BUCAI	P57504 buchera ap
754	6	1.4	655	1	ABG2_HUMAN	Q9unc0 homo sapien	827	6	1.4	802	1	Y250_HUMAN	Q92540 homo sapien
755	6	1.4	656	1	UVRB_MYCCE	P47319 mycoplasma	828	6	1.4	805	1	PRIA_BACSU	P94461 bacillus su
756	6	1.4	657	1	KNOB_PLAFN	P06719 plasmodium	829	6	1.4	810	1	YL25_CAEEL	P34418 caenorhabdi
757	6	1.4	665	1	LAMA_XENLA	P11048 xenopus lae	830	6	1.4	811	1	MUTS_THEAQ	O56215 thermus aqu
758	6	1.4	667	1	Y366_MYCCE	P47606 mycoplasma	831	6	1.4	816	1	HUNB_DROVI	P13361 drosophila
759	6	1.4	669	1	AMY_ALTHA	P29957 alteromonas	832	6	1.4	817	1	MUTS_THECA	Q01778 musca domes
760	6	1.4	671	1	Z282_HUMAN	Q9udv7 homo sapien	833	6	1.4	817	1	HUNB_THETH	O9ziX6 thermus cal
761	6	1.4	671	1	POLG_PRSWV	P19724 papaya ring	834	6	1.4	818	1	MUTS_THETH	Q56239 thermus the
762	6	1.4	677	1	BS4_MOUSE	P54729 mus musculu	835	6	1.4	818	1	YX07_CAEEL	Q11114 caenorhabdi
763	6	1.4	678	1	GARP_PLAFC	P13816 plasmodium	836	6	1.4	819	1	AHM1_ARATH	Q9m3h5 arabidopsis

837	6	1.4	820	1	FES_FELCA	P14238	felis silve	910	1	1012	1	SCA4_RICSL	Q9a180	rickettsia	
838	6	1.4	820	1	FES_MOUSE	P16879	mus musculus	911	6	1.4	1013	1	SCA4_RICRH	Q9a181	rickettsia
839	6	1.4	822	1	CAD3_MOUSE	P10287	mus musculus	912	6	1.4	1018	1	FNBA_STAAU	P14738	staphylococ
840	6	1.4	827	1	CADH_RAT	P55281	rattus norv	913	6	1.4	1018	1	SCA4_RICJA	Q9a179	rickettsia
841	6	1.4	832	1	TRNL_CANAL	P43075	candida alb	914	6	1.4	1018	1	YC14_METJA	Q58611	methanococ
842	6	1.4	835	1	UREA_SCHPO	O00084	schizosacch	915	6	1.4	1046	1	HUUA_HAEIN	Q48153	haemophilus
843	6	1.4	835	1	Y422_MYCGE	P47661	mycoplasma	916	6	1.4	1048	1	P100_HCWA	P08318	human cytom
844	6	1.4	838	1	AXN2_RAT	O70240	rattus norv	917	6	1.4	1050	1	TLR7_MOUSE	P58681	mus musculus
845	6	1.4	843	1	CVAA_HAEIN	P40134	haemophilus	918	6	1.4	1052	1	RPOC_BACAN	P77819	bacillus an
846	6	1.4	843	1	EF2_BETVU	O23755	beta vulgar	919	6	1.4	1057	1	POLR_DROME	P16423	drosophila
847	6	1.4	849	1	PRO5_YEAST	P21372	saccharomyc	920	6	1.4	1057	1	SEF1_YEAST	P34228	saccharomyc
848	6	1.4	854	1	PI4K_HUMAN	P42356	homo sapien	921	6	1.4	1071	1	SEF1_KLULA	P87164	kluveromyc
849	6	1.4	859	1	YDBB_SCHPO	Q10362	schizosacch	922	6	1.4	1075	1	PST2_SCHPO	O13919	schizosacch
850	6	1.4	866	1	MYSP_SCHJA	Q05870	schistosoma	923	6	1.4	1076	1	CARB_ARCFU	Q28994	archaeoglob
851	6	1.4	866	1	MYSP_SCHNA	P06198	schistosoma	924	6	1.4	1080	1	RPOB_MESVI	Q9mus5	mesostigma
852	6	1.4	870	1	YCS0_YEAST	P25623	saccharomyc	925	6	1.4	1081	1	RPOB_CVACA	Q9tm35	cyanidium c
853	6	1.4	871	1	CPG2_MOUSE	Q98833	mus musculus	926	6	1.4	1081	1	SYI_TETH	P36422	tetrahymena
854	6	1.4	878	1	MSH4_YEAST	P40965	saccharomyc	927	6	1.4	1085	1	IFH1_YEAST	P35520	saccharomyc
855	6	1.4	879	1	RA50_PYRHO	O58687	pyrococcus	928	6	1.4	1085	1	RA26_YEAST	P40352	saccharomyc
856	6	1.4	879	1	RA50_SULTO	Q96yrs	sulfolobus	929	6	1.4	1087	1	RP16_HUMAN	Q9ula9	homo sapien
857	6	1.4	880	1	BRC4_DROME	Q24206	drosophila	930	6	1.4	1087	1	RP16_MOUSE	Q9epk7	mus musculus
858	6	1.4	886	1	LEUR_YEAST	P08638	saccharomyc	931	6	1.4	1089	1	Y553_HUMAN	Q9ukj3	homo sapien
859	6	1.4	893	1	Y492_CAEEL	P34531	caenorhabdi	932	6	1.4	1092	1	DHE2_YEAST	P33327	saccharomyc
860	6	1.4	895	1	Z2B1_HUMAN	Q9y2x9	homo sapien	933	6	1.4	1099	1	CARB_THEMA	Q9wz27	thermotoga
861	6	1.4	896	1	CB1_MOUSE	P22682	mus musculus	934	6	1.4	1100	1	JAK3_RAT	Q63272	rattus norv
862	6	1.4	898	1	KBF2_HUMAN	Q00653	homo sapien	935	6	1.4	1101	1	LOS1_YEAST	P33418	saccharomyc
863	6	1.4	898	1	NIA_BETVE	P27783	betula verr	936	6	1.4	1101	1	FURC_DROME	P30430	drosophila
864	6	1.4	899	1	V120_HSVSA	Q01055	herpesvirus	937	6	1.4	1108	1	DBS_HUMAN	O15068	homo sapien
865	6	1.4	905	1	RP11_THECE	P31813	thermococcu	938	6	1.4	1124	1	JAK3_HUMAN	P52333	homo sapien
866	6	1.4	905	1	YAG1_SCHPO	Q09866	schizosacch	939	6	1.4	1125	1	MFD_BORBU	O51568	borrelia bu
867	6	1.4	906	1	CB1_HUMAN	P22681	homo sapien	940	6	1.4	1131	1	RPOB_ANASP	P22703	anabaena sp
868	6	1.4	908	1	H104_YEAST	P31539	saccharomyc	941	6	1.4	1137	1	A4E1_HUMAN	Q9upm8	homo sapien
869	6	1.4	909	1	CNG4_HUMAN	Q14028	homo sapien	942	6	1.4	1142	1	GIN4_YEAST	Q12263	saccharomyc
870	6	1.4	915	1	LIQ_DROME	Q24118	drosophila	943	6	1.4	1152	1	YBF7_CAEEL	Q21286	caenorhabdi
871	6	1.4	916	1	TOP1_ARATH	P30181	arabidopsis	944	6	1.4	1159	1	N124_SCHPO	Q09904	schizosacch
872	6	1.4	919	1	PWP2_HUMAN	Q15269	homo sapien	945	6	1.4	1159	1	RPOC_FORCN	O33431	porphyromon
873	6	1.4	919	1	SVI1_THEMA	P46213	thermotoga	946	6	1.4	1172	1	AHM2_ARATH	O64474	arabidopsis
874	6	1.4	921	1	DPOL1_RICHE	Q9rib6	rickettsia	947	6	1.4	1172	1	TRP2_MOUSE	Q9r244	mus musculus
875	6	1.4	922	1	DPOL1_RICFE	Q9raa9	rickettsia	948	6	1.4	1174	1	NOS_RHOPR	Q26240	rhodnius pr
876	6	1.4	925	1	PDC2_YEAST	P32896	saccharomyc	949	6	1.4	1205	1	ATS3_HUMAN	O15072	homo sapien
877	6	1.4	926	1	POOL_HAEIN	P45181	haemophilus	950	6	1.4	1211	1	BUN2_DROME	Q24523	drosophila
878	6	1.4	928	1	RB_HUMAN	O06400	homo sapien	951	6	1.4	1215	1	BGCN_DROME	Q9w112	drosophila
879	6	1.4	930	1	SVI1_STRPN	Q92hb3	streptococc	952	6	1.4	1220	1	IF2P_HUMAN	O60841	homo sapien
880	6	1.4	933	1	CG54_YEAST	P30665	saccharomyc	953	6	1.4	1224	1	RPOD_PINTH	P41606	pinus thunb
881	6	1.4	934	1	CITC_HUMAN	P11586	h c-1-tetra	954	6	1.4	1232	1	ADDA_BACSU	P23478	bacillus su
882	6	1.4	941	1	GBR2_HUMAN	Q75899	homo sapien	955	6	1.4	1233	1	YF16_YEAST	P43597	saccharomyc
883	6	1.4	943	1	ROR2_HUMAN	Q01974	homo sapien	956	6	1.4	1234	1	CPAH_MOUSE	P06909	mus musculus
884	6	1.4	944	1	VGLB_HSVT2	Q9wrl5	herpesvirus	957	6	1.4	1251	1	RBP2_PLAVB	Q00799	plasmodium
885	6	1.4	946	1	YBT6_YEAST	P38250	saccharomyc	958	6	1.4	1269	1	FURL_DROME	P26016	drosophila
886	6	1.4	950	1	URB1_USTWA	P40349	ustilago ma	959	6	1.4	1286	1	PATC_DROME	P18502	drosophila
887	6	1.4	953	1	PSD1_HUMAN	Q99460	homo sapien	960	6	1.4	1287	1	VAC2_HELPY	Q48245	helicobacte
888	6	1.4	953	1	PSD1_RAT	O88761	rattus norv	961	6	1.4	1288	1	VACA_HELPJ	Q924w5	helicobacte
889	6	1.4	954	1	B1R1_YEAST	P47134	saccharomyc	962	6	1.4	1290	1	PIG1_HUMAN	P19174	homo sapien
890	6	1.4	956	1	PHK3_NICPL	Q08436	nicotiana p	963	6	1.4	1290	1	PIG1_RAT	P10886	rattus norv
891	6	1.4	956	1	TSP3_HUMAN	P49746	homo sapien	964	6	1.4	1290	1	VACA_HELPY	P55981	helicobacte
892	6	1.4	956	1	TSP3_MOUSE	Q05895	mus musculus	965	6	1.4	1291	1	PIG1_BOVIN	P08487	bos taurus
893	6	1.4	958	1	BP28_MACFA	Q9gm44	macaca fasc	966	6	1.4	1291	1	VAC4_HELPY	Q48258	helicobacte
894	6	1.4	960	1	BPD1_MOUSE	P52734	mus musculus	967	6	1.4	1296	1	VAC1_HELPY	Q48247	helicobacte
895	6	1.4	961	1	FGD1_HUMAN	P98174	homo sapien	968	6	1.4	1299	1	JAK3_MOUSE	Q62137	mus musculus
896	6	1.4	967	1	YSGO_CLOAB	P33747	clostridium	969	6	1.4	1310	1	VAC3_HELPY	Q48253	helicobacte
897	6	1.4	975	1	CITM_YEAST	P09440	s c-1-tetra	970	6	1.4	1312	1	PIF1_DROME	P25455	drosophila
898	6	1.4	977	1	DLP3_RAT	Q97838	rattus norv	971	6	1.4	1325	1	G160_MOUSE	P55937	mus musculus
899	6	1.4	979	1	P115_MYCHR	P41508	mycoplasma	972	6	1.4	1327	1	TNK1_HUMAN	Q95271	homo sapien
900	6	1.4	979	1	PTPN_BOVIN	P56722	bos taurus	973	6	1.4	1339	1	DPQA_LEIDO	O00874	leishmania
901	6	1.4	979	1	PTPN_MOUSE	Q60673	mus musculus	974	6	1.4	1359	1	STH1_YEAST	P32597	saccharomyc
902	6	1.4	988	1	OMB_DROME	Q24432	drosophila	975	6	1.4	1362	1	BRD4_HUMAN	O60885	homo sapien
903	6	1.4	991	1	SCA4_RICSI	Q9aj77	rickettsia	976	6	1.4	1367	1	AMVH_YEAST	P08640	saccharomyc
904	6	1.4	993	1	SCP1_MOUSE	Q62209	mus musculus	977	6	1.4	1374	1	YC9A_SCHPO	Q08884	schizosacch
905	6	1.4	997	1	SCP1_RAT	O03410	rattus norv	978	6	1.4	1395	1	SP41_YEAST	P38904	saccharomyc
906	6	1.4	999	1	M2B1_BOVIN	Q29451	bos taurus	979	6	1.4	1410	1	PDRB_YEAST	P40550	saccharomyc
907	6	1.4	1010	1	SCM4_RICPA	Q9aj75	rickettsia	980	6	1.4	1420	1	YM8B_YEAST	Q03496	saccharomyc
908	6	1.4	1011	1	SCA4_RICAF	Q9aj83	rickettsia	981	6	1.4	1423	1	FRUA_STRMU	Q03174	streptococ
909	6	1.4	1011	1	SCA4_RICMN	Q9aj82	rickettsia	982	6	1.4	1425	1	MADI_HUMAN	Q95405	homo sapien

883 6 1.4 1426 1 EGFR_DROME P04412 drosophila
 884 6 1.4 1431 1 DAPK_HUMAN P53355 homo sapien
 885 6 1.4 1432 1 SKI3_YEAST P17883 saccharomyc
 886 6 1.4 1447 1 SGL1_YEAST P35187 saccharomyc
 887 6 1.4 1459 1 GEA2_YEAST P39993 saccharomyc
 888 6 1.4 1464 1 NCO2_HUMAN Q15596 homo sapien
 889 6 1.4 1466 1 SPA2_YEAST P23201 saccharomyc
 890 6 1.4 1473 1 NAL1_HUMAN Q9C000 homo sapien
 891 6 1.4 1483 1 UFD4_YEAST P33202 saccharomyc
 892 6 1.4 1489 1 YGP0_YEAST P53115 saccharomyc
 893 6 1.4 1504 1 DPO2_YEAST P14284 saccharomyc
 894 6 1.4 1517 1 GLSF_SPIOL Q43155 spinacia ol
 895 6 1.4 1533 1 PUM1_DROME P25822 drosophila
 896 6 1.4 1581 1 ARO1_PNECA Q12659 p pentafunc
 897 6 1.4 1586 1 SNR21_HUMAN P51531 homo sapien
 898 6 1.4 1603 1 VIT5_CAEEL P06125 caenorhabdi
 899 6 1.4 1616 1 APXL_HUMAN Q13796 homo sapien
 1000 6 1.4 1629 1 RRPO_SHMV P89202 sunn-hemp m

ALIGNMENTS

RESULT 1
 MSPL_PLAFPC STANDARD; PRT; 1726 AA.
 AC P04934;
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMMSA) (P195).
 DE MSP-1.
 GN Plasmodium falciparum (isolate Camp / Malaysia).
 OS Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 OX NCBI_TaxID=5835;
 RN [1]
 RP SEQUENCE OF 1104-1726 FROM N.A.
 RX MEDLINE=86205236; PubMed=3517809;
 RA Weber J.L., Leininger W.M., Lyon J.A.;
 RT "Variation in the gene encoding a major merozoite surface antigen of
 the human malaria parasite Plasmodium falciparum.";
 RL Nucleic Acids Res. 14:3311-3323(1986).
 RN [2]
 RP SEQUENCE OF 1104-1726 FROM N.A.
 RX MEDLINE=88143999; PubMed=3278296;
 RA Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
 RT "Merozoite surface protein sequence from the Camp strain of the human
 malaria parasite Plasmodium falciparum.";
 RL Nucleic Acids Res. 16:1206-1206(1988).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X03831; CAA27446.1;
 DR PTR; A23386; SAZQGM.
 DR InterPro; IPR000561; EGF-like.
 DR Pfam; PF00008; EGF; 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1726 AA; 196197 MW; DB8AD45FA352BCF3 CRC64;
 Query Match 53.6%; Score 231; DB 1; Length 1726;
 Best Local Similarity 100.0%; Pred. No. 3.5e-226;
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISVTMDNLSGFENEDVYILKPLAGVYRSLLKQIKNIETFNLNLDIINSRLKRRY 117
 |||||
 DB 1333 AISVTMDNLSGFENEDVYILKPLAGVYRSLLKQIKNIETFNLNLDIINSRLKRRY 1392
 |||||
 QY 118 FLDVLESDLMQFKHISSEYIIESEKLLNSEQNTLLKSYKIKESVENDIKFAEGIS 177
 |||||
 DB 1393 FLDVLESDLMQFKHISSEYIIESEKLLNSEQNTLLKSYKIKESVENDIKFAEGIS 1452
 |||||
 QY 178 YEKVLAKYKDDLESIRKVIKEEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 237
 |||||
 DB 1453 YEKVLAKYKDDLESIRKVIKEEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 1512
 |||||
 QY 238 YNNLVNKIDDYILNLIKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKN 288
 |||||
 DB 1513 YNNLVNKIDDYILNLIKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKN 1563

RESULT 2

MSPL_PLAFPC STANDARD; PRT; 1726 AA.
 AC P50495;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMMSA) (GP195).
 GN MSP-1.
 OS Plasmodium falciparum (isolate Palo Alto / Uganda).
 OX Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 OX NCBI_TaxID=57270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89005525; PubMed=3049134;
 RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
 RA Siddiqui W.A.;
 RT "Plasmodium falciparum: gene structure and hydrophobic profile of the
 major merozoite surface antigen (gp195) of the Uganda-Palo Alto
 isolate.";
 RL Exp. Parasitol. 67:1-11(1988).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M37213; AAA29611.1;
 DR InterPro; IPR000561; EGF-like.

DR Pfam: PF00008; EGF; 1.
KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
FT Transmembrane; GPI-anchor.
FT SIGNAL 1 19
FT CHAIN 20 1726
FT CARBOHYD 133 1726
FT CARBOHYD 133 1726
FT CARBOHYD 272 1726
FT CARBOHYD 501 1726
FT CARBOHYD 567 1726
FT CARBOHYD 638 1726
FT CARBOHYD 827 1726
FT CARBOHYD 924 1726
FT CARBOHYD 944 1726
FT CARBOHYD 990 1726
FT CARBOHYD 1016 1726
FT CARBOHYD 1114 1726
FT CARBOHYD 1221 1726
FT CARBOHYD 1613 1726
SQ SEQUENCE 1726 AA; 5859CEFA2F9A026 CRC64;
Query Match 53.6%; Score 231; DB 1; Length 1726;
Best Local Similarity 100.0%; Pred. No. 3.5e-226;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 58 AISTVMDNLISGFENEYDVIYKPLAGVYRSLLKQTEKIFTNLNLNLSRLKRRY 1392
DB 1333 AISTVMDNLISGFENEYDVIYKPLAGVYRSLLKQTEKIFTNLNLNLSRLKRRY 1392
QY 118 FLDVLESMDQPKHSSNYYIEDSKLLNSQKNTLLSKYKVESVNDIKFAQEGIS 177
DB 1393 FLDVLESMDQPKHSSNYYIEDSKLLNSQKNTLLSKYKVESVNDIKFAQEGIS 1452
QY 178 YVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAKTDEQKESKFLPLTNIETL 237
DB 1453 YVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAKTDEQKESKFLPLTNIETL 1512
QY 238 YNNLVNKIDDYLNKAKINDCNVDEAHVKITKLSDLKAIDDKIDLFKN 288
DB 1513 YNNLVNKIDDYLNKAKINDCNVDEAHVKITKLSDLKAIDDKIDLFKN 1563
RESULT 3
MSPI_PLAFF STANDARD; PRT; 1701 AA.
AC P13819;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMSA).
GN MSP-1.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5837;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88142999; PubMed=2449612;
RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
RA Brown G.V., Anders R.F., Kemp D.J.;
RT "Variation in the precursor to the major merozoite surface antigens
RT of Plasmodium falciparum";
RL Mol. Biochem. Parasitol. 27:291-302(1988).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M19143; AA29653.1; -.
CC PIR: A54498; A54498.
CC InterPro: IPR000561; EGF-like.
CC Pfam: PF00008; EGF; 1.
KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
FT Transmembrane; GPI-anchor.
FT SIGNAL 1 19
FT CHAIN 20 1701
FT CARBOHYD 110 1701
FT CARBOHYD 239 1701
FT CARBOHYD 470 1701
FT CARBOHYD 536 1701
FT CARBOHYD 607 1701
FT CARBOHYD 802 1701
FT CARBOHYD 899 1701
FT CARBOHYD 919 1701
FT CARBOHYD 965 1701
FT CARBOHYD 1089 1701
FT CARBOHYD 1196 1701
FT CARBOHYD 1588 1701
SQ SEQUENCE 1701 AA; 3920B75E73D38552 CRC64;
Query Match 44.1%; Score 190; DB 1; Length 1701;
Best Local Similarity 100.0%; Pred. No. 1.6e-184;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 99 TFNLNLDILNLSRLKRRYFLOVLESMDLQKHISSNYYIEDSKLLNSQKNTLLSKY 158
DB 1349 TFNLNLDILNLSRLKRRYFLOVLESMDLQKHISSNYYIEDSKLLNSQKNTLLSKY 1408
QY 159 KYIKESVNDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAKT 218
DB 1409 KYIKESVNDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAKT 1468
QY 219 DEQKESKFLPLTNIETLYNNLVNKIDDYLNKAKINDCNVDEAHVKITKLSDLKA 278
DB 1469 DEQKESKFLPLTNIETLYNNLVNKIDDYLNKAKINDCNVDEAHVKITKLSDLKA 1528
QY 279 IDDKIDLFKN 288
DB 1529 IDDKIDLFKN 1538
RESULT 4
MSPI_PLAFF STANDARD; PRT; 1701 AA.
AC P08569;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=70153;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88011243; PubMed=3079521;
RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT "Allelic dimorphism in a surface antigen gene of the malaria parasite
RT Plasmodium falciparum";
RL J. Mol. Biol. 195:273-287(1987).
RN [2]
RP REVISIONS TO 1403; 1569 AND 1629.
RA Tanabe K.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-115 FROM N.A.

```

RX MEDLINE-86136024; PubMed-3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA Stunnenberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of
RT Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05624; CAA29112.1; -
DR PIR; A26868; A26868.
DR PIR; B25120; B25120.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96EA98 CRC64;

Query Match 44.18; Score 190; DB 1; Length 1701;
Best Local Similarity 100.0%; Pred. No. 1.6e-184;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 TFNLNLDILNSRLKRRKYFDVLESQKFISSNEYIIEFSKLLNSQKNTLKSY 158
Db 1349 TFNLNLDILNSRLKRRKYFDVLESQKFISSNEYIIEFSKLLNSQKNTLKSY 1408

Qy 159 KYIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAKT 218
Db 1409 KYIKESVENDIKFAQEGISYYEKVLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAKT 1468

Qy 219 DEQKESKFLPFLNFIETLYNNLVNKIDYLYNLKAKINDCNVKEDEAHVKITKLSDLKA 278
Db 1469 DEQKESKFLPFLNFIETLYNNLVNKIDYLYNLKAKINDCNVKEDEAHVKITKLSDLKA 1528

Qy 279 IDDKIDLFKN 288
Db 1529 IDDKIDLFKN 1538

RESULT 5
MSPL_PLAF3
ID MSPL_PLAF3 STANDARD; PRT; 1682 AA.
AC P19598; Q25921;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)

```

```

DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate ro-33 / Ghana).
OX Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5834;
RN [1]
RP SEQUENCE OF 1-1061 FROM N.A.
RX MEDLINE-88166657; PubMed-3327688;
RA Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT "A naturally occurring gene encoding the major surface antigen
RT precursor p190 of Plasmodium falciparum lacks tripeptide repeats.";
RL EMBO J. 6:4137-4142(1987).
RN [2]
RP SEQUENCE OF 1032-1682 FROM N.A.
RX MEDLINE-95354793; PubMed-7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
RT merozoite surface antigen-1.";
RL Exp. Parasitol. 81:47-54(1995).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M35727; AAA29715.1; -
DR EMBL; Y00087; CAA68280.1; -
DR EMBL; Z35326; CAA84555.1; -
DR PIR; S06286; S06286.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1682 MEROZOITE SURFACE PROTEIN 1.
FT TRANSMEM 1666 1682 MEMBRANE ANCHOR (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 785 785 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 881 881 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 947 947 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1178 1178 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1569 1569 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1682 AA; 192462 MW; CB2A1E159948CAD6 CRC64;

Query Match 39.9%; Score 172; DB 1; Length 1682;
Best Local Similarity 99.7%; Pred. No. 3.2e-166;
Matches 292; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 58 AISVTMDNLSGFFENYDVLYLKLPLAGVYRSLKKQIKBNFTFNLNLDILNSRLKRRKY 117
Db 1290 AISVTMDNLSGFFENYDVLYLKLPLAGVYRSLKKQIKBNFTFNLNLDILNSRLKRRKY 1349

Qy 118 FLDVLESQKFISSNEYIIEFSKLLNSQKNTLKSYIKESVENDIKFAQEGIS 177
Db 1350 FLDVLESQKFISSNEYIIEFSKLLNSQKNTLKSYIKESVENDIKFAQEGIS 1409

Qy 178 YEKVLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAKTDEQKESKFLPFLNFIETL 237
Db 1410 YEKVLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAKTDEQKESKFLPFLNFIETL 1468

```



```

Best Local Similarity 100.0%; Pred. No. 3.8e-38;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 NSGCFRHLDERECKCLLNKQEGDKVCVENPNTCENNNGGCDADA 397
|||||
Db 1540 NSGCFRHLDERECKCLLNKQEGDKVCVENPNTCENNNGGCDADA 1585

RESULT 8
RNP_BOSTR
ID RNP_BOSTR STANDARD; PRT; 124 AA.
AC P07849;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Boselaphus tragocamelus (Nilgai).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Boselaphus.
OX NCBI_TaxID=9917;
RN [1]
RP SEQUENCE.
RC TISSUE-Pancreas;
RX MEDLINE=80109825; PubMed=7353035;
RA Beintema J.J.;
RT "Primary structures of pancreatic ribonucleases from Bovidae. Impala,
RT Thomson's gazelle, nilgai and water buffalo.";
RL Biochim. Biophys. Acta 621:89-102(1980).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
DR InterPro: IPR001427; RNaseA.
DR Pfam: PF00074; RNaseA; 1.
DR PRINTS: PR00794; RIBONUCLEASE.
DR PRODOM: PD000535; RNaseA; 1.
DR SMART: SM00092; RNase_Pc; 1.
DR PROSITE: PS00127; RNASE_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease.
FT DISULFID 26 84
FT DISULFID 40 95
FT DISULFID 58 110
FT DISULFID 65 72
FT ACT_SITE 12 12
FT ACT_SITE 41 41
FT ACT_SITE 119 119
SQ SEQUENCE 124 AA; 13685 MW; 09E84341868BC94C CRC64;

Query Match 3.5%; Score 15; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RNP_TRAOR
ID RNP_TRAOR STANDARD; PRT; 124 AA.
AC P00658;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Tragelaphus oryx (Eland) (Taurotragus oryx).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Tragelaphus.
OX NCBI_TaxID=9945;
RN [1]
RP PARTIAL SEQUENCE.
RC TISSUE-Pancreas;
RX MEDLINE=76184786; PubMed=1268225;
RA Russchen F., de Vrieze G., Gastra W., Beintema J.J.;
RT "Studies on the covalent structure of eland pancreatic ribonuclease.";
RL Biochim. Biophys. Acta 427:719-726(1976).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
DR InterPro: IPR001427; RNaseA.
DR Pfam: PF00074; RNaseA; 1.
DR PRINTS: PR00794; RIBONUCLEASE.
DR PRODOM: PD000535; RNaseA; 1.
DR SMART: SM00092; RNase_Pc; 1.
DR PROSITE: PS00127; RNASE_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease.
FT DISULFID 26 84
FT DISULFID 40 95
FT DISULFID 58 110
FT DISULFID 65 72
FT ACT_SITE 12 12
FT ACT_SITE 41 41
FT ACT_SITE 119 119
SQ SEQUENCE 124 AA; 13742 MW; E612512E71A5D00A CRC64;

Query Match 3.5%; Score 15; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 KETAAKFERQHMD 33
|||||
Db 1 KETAAKFERQHMD 15

RESULT 10
RNP_BOVIN
ID RNP_BOVIN STANDARD; PRT; 150 AA.
AC P00656;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic precursor (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Bos taurus (Bovine), and
OS Bison bison (American bison).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913, 9901;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Bovine;
RX MEDLINE=86262557; PubMed=2838818;
RA Carsana A., Confalone E., Palmieri M., Libonati M., Furia A.;
RT "Structure of the bovine pancreatic ribonuclease gene: the unique
RT intervening sequence in the 5' untranslated region contains a
RT promoter-like element.";
RL Nucleic Acids Res. 16:5491-5502(1988).
RN [2]
RP SEQUENCE OF 23-150 FROM N.A.
RC SPECIES=Bovine;
RX MEDLINE=9603920; PubMed=7479688;
RA Delcardayre S.B., Ribo M., Yokel E.M., Quirk D.J., Rutter W.J.,

```

RA Raines R.T.;
RT "Engineering ribonuclease A: production, purification and
RL characterization of wild-type enzyme and mutants at Gln11.";
RN Protein Eng. 8:261-273(1995).
RX [3]
RP SEQUENCE OF 27-150, AND DISULFIDE BONDS.
RC SPECIES=Bovine;
RA Smyth D.G., Stein W.H., Moore S.;
RT "The sequence of amino acid residues in bovine pancreatic
RL ribonuclease: revisions and confirmations.";
RN J. Biol. Chem. 238:227-234(1963).
RX [4]
RP SEQUENCE OF 27-150.
RC SPECIES=Bovine;
RA Plummer T.H. Jr., Hirs C.H.W.;
RT "On the structure of bovine pancreatic ribonuclease B. Isolation of
RL a glycopeptide.";
RN J. Biol. Chem. 239:2530-2538(1964).
RX [5]
RP ACTIVE SITE.
RC SPECIES=Bovine;
RA Heinrichson R.L., Stein W.H., Crestfield A.M., Moore S.;
RT "The reactivities of the histidine residues at the active site of
RL ribonuclease toward halo acids of different structures.";
RN J. Biol. Chem. 240:2921-2934(1965).
RX [6]
RP ACTIVE SITE.
RC SPECIES=Bovine;
RA Shall S., Barnard E.A.;
RT "Heavy atom-labelled derivatives of bovine pancreatic ribonuclease.
RL I. Specific reactions of ribonuclease with N-acetylhomocysteine
thiolactone and silver ion.";
RN J. Mol. Biol. 41:237-251(1969).
RX [7]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RC SPECIES=Bovine;
RA Wyckoff H.W., Tsernoglou D., Hanson A.W., Knox J.R., Lee B.,
RT Richards F.M.;
RL "The three-dimensional structure of ribonuclease-S. Interpretation of
an electron density map at a nominal resolution of 2 A.";
RN J. Biol. Chem. 245:305-328(1970).
RX [8]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RC SPECIES=Bovine;
RA Carlisle C.H., Palmer R.A., Mazumdar S.K., Gorinsky B.A.,
RT Yeates D.G.R.;
RL "The structure of ribonuclease at 2.5-A resolution.";
RN J. Mol. Biol. 85:1-18(1974).
RX [9]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RC SPECIES=Bovine;
RA Wlodawer A., Bott R., Sjoelin L.;
RT "The refined crystal structure of ribonuclease A at 2.0-A
resolution.";
RN J. Biol. Chem. 257:1325-1332(1982).
RX [10]
RP X-RAY CRYSTALLOGRAPHY (1.26 ANGSTROMS).
RC SPECIES=Bovine;
RA Wlodawer A., Svensson L.A., Sjoelin L., Gilliland G.L.;
RT "Structure of phosphate-free ribonuclease A refined at 1.26 A.";
RN Biochemistry 27:2705-2717(1988).
RX [11]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RC SPECIES=Bovine;
RA Leonidas D.D., Shapiro R., Irons L.I., Russo N., Acharya K.R.;
RT "Crystal structures of ribonuclease A complexes with 5'-
diphosphoadenosine 3'-phosphate and 5'-diphosphoadenosine
2'-phosphate at 1.7-A resolution.";

RL Biochemistry 36:5578-5588(1997).
RN [12]
RP STRUCTURE BY NMR.
RC SPECIES=Bovine;
RX MEDLINE=89375325; PubMed=2775743;
RA Robertson A.D., Purisma E.O., Eastman M.A., Scheraga H.A.;
RT "Proton NMR assignments and regular backbone structure of bovine
RL pancreatic ribonuclease A in aqueous solution.";
RN Biochemistry 28:5930-5938(1989).
RX [13]
RP STRUCTURE BY NMR.
RC SPECIES=Bovine;
RX MEDLINE=89377830; PubMed=2776756;
RA Rico M., Bruix M., Santoro J., Gonzalez C., Neira J.L., Nieto J.L.,
RT Herranz J.;
RL "Sequential 1H-NMR assignment and solution structure of bovine
pancreatic ribonuclease A.";
RN Eur. J. Biochem. 183:623-638(1989).
RX [14]
RP STRUCTURE BY NMR.
RC SPECIES=Bovine;
RX MEDLINE=93044359; PubMed=1841699;
RA Rico M., Santoro J., Gonzalez C., Bruix M., Neira J.L., Nieto J.L.,
RT Herranz J.;
RL "3D structure of bovine pancreatic ribonuclease A in aqueous
solution: an approach to tertiary structure determination from a
small basis of 1H NMR NOE correlations.";
RN J. Biomol. NMR 1:283-298(1991).
RX [15]
RP DNA-BINDING.
RC SPECIES=Bovine;
RX MEDLINE=86179900; PubMed=3961503;
RA McPherson A., Brayer G., Cascio D., Williams R.;
RT "The mechanism of binding of a polynucleotide chain to pancreatic
RL ribonuclease.";
RN Science 232:765-768(1986).
RX [16]
RP SEQUENCE OF 27-150.
RC SPECIES=B. bison;
RX MEDLINE=76259396; PubMed=955781;
RA Muskiet F.A.J., Welling G.W., Beintema J.J.;
RT "Studies on the primary structure of bison pancreatic ribonuclease.";
RL Int. J. Pept. Protein Res. 8:345-348(1976).
RX [17]
RP CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
with 2',3'-cyclic phosphate intermediates.
RN [18]
RP SUBCELLULAR LOCATION: Secreted.
RX [19]
RP TISSUE SPECIFICITY: PANCREAS.
RN [20]
RP MISCELLANEOUS: RIBONUCLEASE CAN DESTABILIZE OR UNWIND THE DNA
HELIX BY COMPLEXING WITH SINGLE-STRANDED DNA; THIS COMPLEX ARISES
BY AN EXTENDED MULTISITE CATION-ANTION INTERACTION BETWEEN THE
LYSINE AND ARGININE RESIDUES OF THE ENZYME AND THE PHOSPHATE
GROUPS OF THE NUCLEOTIDES.
RN [21]
RP MISCELLANEOUS: THE BISON SEQUENCE APPEARS TO BE IDENTICAL WITH
THAT OF BOVINE.
RX [22]
RP SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
RN [23]
RP DATABASE: NAME=Worthington enzyme manual;
WWW="http://www.worthington-biochem.com/manual/R/RNASE.html".
RX [24]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
RN [25]
RP EMBL: X07283; CAA30263.1;
RX EMBL: S80747; AAB35594.1;
RA PIR: A00804; NRBO.
RT PIR: A91771; NRBO.
RL PIR: A32471; A32471.
RN PIR: S00897; S00897.


```

FT  NON_TER      30
SQ  SEQUENCE      30 AA; 3374 MW; 9184E7C816394D20 CRC64;

Query Match      3.2%; Score 14; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.le-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  19 KETAAKFERQHMD 32
Db  1 KETAAKFERQHMD 14
    |||||
    |||||

RESULT 12
RNP_AEPME
ID  _RNP_AEPME      STANDARD;      PRT;      124 AA.
AC  P07847;
DT  01-AUG-1988 (Rel. 08, Created)
DT  01-AUG-1988 (Rel. 08, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Ribonuclease pancreatic (RC 3.1.27.5) (RNase 1) (RNase A).
GN  RNASE1 OR RNS1.
OS  Apeyceros melampus (Impala).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC  Bovidae; Apeycerotinae; Apeyceros.
OC  NCBI_TaxID=9897;
RN  [1]
RP  SEQUENCE.
RC  TISSUE=Pancreas;
RX  MEDLINE=80109825; PubMed=7353035;
RA  Beintema J.J.;
RT  "Primary structures of pancreatic ribonucleases from Bovidae. Impala,
RT  Thomson's gazelle, nilgai and water buffalo.";
RL  Biochim. Biophys. Acta 621:89-102(1980).
CC  -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC  phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC  with 2',3'-cyclic phosphate intermediates.
CC  -1- SUBCELLULAR LOCATION: Secreted.
CC  -1- TISSUE SPECIFICITY: PANCREAS.
CC  -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
DR  PIR; S07141; S07141.
DR  HSSP; P00656; 1SRN.
DR  InterPro; IPR001427; RNaseA.
DR  Pfam; PF00074; rnasea; 1.
DR  PRINTS; PR00794; RIBONUCLEASE.
DR  PRODOM; PD000535; RNasea; 1.
DR  SMART; SM00092; RNase_Pc; 1.
DR  PROSITE; PS00127; RNASE_PANCREATIC; 1.
KW  Hydrolase; Nuclease; Endonuclease; Glycoprotein.
FT  CARBOHYD 34 N-LINKED (GLCNAC. .) (PARTIAL).
FT  DISULFID 26 84 BY SIMILARITY.
FT  DISULFID 40 95 BY SIMILARITY.
FT  DISULFID 58 110 BY SIMILARITY.
FT  DISULFID 65 72 BY SIMILARITY.
FT  ACT_SITE 12 12 BY SIMILARITY.
FT  ACT_SITE 41 41 BY SIMILARITY.
FT  ACT_SITE 119 119 BY SIMILARITY.
SQ  SEQUENCE 124 AA; 13678 MW; B3018B023584CF5 CRC64;

Query Match      2.8%; Score 12; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  22 AAARKFERQHMDS 33
Db  4 AAARKFERQHMDS 15
    |||||
    |||||

RESULT 13
RNP_ANTAM
ID  _RNP_ANTAM      STANDARD;      PRT;      124 AA.
AC  P00668;
DT  21-JUL-1986 (Rel. 01, Created)

```

DT 21-JUL-1986 (Rel. 01, Last sequence update) .
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
 GN RNASE1 OR RNS1.
 OS Antilocapra americana (Pronghorn).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Antilocapridae; Antilocapra.
 OX NCBI_TaxID=9891;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Pancreas;
 RX MEDLINE=80075014; PubMed=513141;
 RA Beintema J.J., Gastra W., Munniksmas J.;
 RT "Primary structure of pronghorn pancreatic ribonuclease: close
 relationship between giraffe and pronghorn.";
 RL J. Mol. Evol. 13:305-316(1979).
 CC -|- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
 phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
 with 2',3'-cyclic phosphate intermediates.
 CC -|- SUBCELLULAR LOCATION: Secreted.
 CC -|- TISSUE SPECIFICITY: PANCREAS.
 CC -|- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
 DR PIR: A00813; NRPB.
 DR HSP: P00656; IRBG.
 DR InterPro: IPR001427; RNaseA.
 DR Pfam: PF00074; rnaase; 1.
 DR PRINTS: PR00794; RIBONUCLEASE.
 DR PRODOM: PD000535; RNaseA; 1.
 DR SMART: SM00092; RNase_Pc; 1.
 DR PROSITE: PS00127; RNASE_PANCREATIC; 1.
 KW Hydrolyase; Nuclease; Endonuclease; Glycoprotein.
 FT DISULFID 26 84 BY SIMILARITY.
 FT DISULFID 40 95 BY SIMILARITY.
 FT DISULFID 58 110 BY SIMILARITY.
 FT DISULFID 65 72 BY SIMILARITY.
 FT ACT_SITE 12 12 BY SIMILARITY.
 FT ACT_SITE 41 41 BY SIMILARITY.
 FT ACT_SITE 119 119 BY SIMILARITY.
 FT CARBOHYD 34 34 N-LINKED (GLCNAC...) (PARTIAL).
 SQ SEQUENCE 124 AA: 13711 MW: 9435EF532420F852 CRC64;

Query Match 2.8%; Score 12; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.00012;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 KETAAKFERQH 30
 |||||
 DB 1 KETAAKFERQH 12

RESULT 14
 RNP_CONTA
 ID RNP_CONTA STANDARD; PRT; 124 AA.
 AC P00660;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
 GN RNASE1 OR RNS1.
 OS Connochaetes taurinus (Brindled gnu).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Alcelaphinae; Connochaetes.
 OX NCBI_TaxID=9927;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Pancreas;
 RX MEDLINE=76210818; PubMed=1227969;
 RA Groen G., Wellington G.W., Beintema J.J.;
 RT "The amino acid sequence of gnu pancreatic ribonuclease.";
 RL FEBS Lett. 60:300-304(1975).
 CC -|- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-

phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
 with 2',3'-cyclic phosphate intermediates.
 CC -|- SUBCELLULAR LOCATION: Secreted.
 CC -|- TISSUE SPECIFICITY: PANCREAS.
 CC -|- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
 DR PIR: A00808; NRGN.
 DR HSP: P00656; ISRN.
 DR InterPro: IPR001427; RNaseA.
 DR Pfam: PF00074; rnaase; 1.
 DR PRINTS: PR00794; RIBONUCLEASE.
 DR PRODOM: PD000535; RNaseA; 1.
 DR SMART: SM00092; RNase_Pc; 1.
 DR PROSITE: PS00127; RNASE_PANCREATIC; 1.
 KW Hydrolyase; Nuclease; Endonuclease.
 FT DISULFID 26 84
 FT DISULFID 40 95
 FT DISULFID 58 110
 FT DISULFID 65 72
 FT ACT_SITE 12 12
 FT ACT_SITE 41 41
 FT ACT_SITE 119 119
 SQ SEQUENCE 124 AA: 13686 MW: 3564BE9D9D871195 CRC64;
 Query Match 2.8%; Score 12; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.00012;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 AAAKFERQHMD 33
 |||||
 DB 4 AAAKFERQHMD 15

RESULT 15
 RNP_SHEEP
 ID RNP_SHEEP STANDARD; PRT; 124 AA.
 AC P00661; P04420;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
 GN RNASE1 OR RNS1.
 OS Ovis aries (Sheep), and
 OS Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 OX NCBI_TaxID=9940, 9925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Sheep, and C.hircus;
 RX MEDLINE=96139017; PubMed=8587129;
 RA Confalone E., Beintema J.J., Sasso M.P., Carsana A., Palmieri M.,
 RA Vento M.T., Furia A.;
 RT "Molecular evolution of genes encoding ribonucleases in ruminant
 species.";
 RL J. Mol. Evol. 41:850-858(1995).
 RN [2]
 RP SEQUENCE.
 RC SPECIES=Sheep, and C.hircus; TISSUE=Pancreas;
 RX MEDLINE=74309062; PubMed=4855010;
 RA Wellington G.W., Scheffer A.J., Beintema J.J.;
 RT "The primary structure of goat and sheep pancreatic ribonucleases.";
 RL FEBS Lett. 41:58-61(1974).
 RN [3]
 RP SEQUENCE.
 RC SPECIES=Sheep;
 RX MEDLINE=74025520; PubMed=4356260;
 RA Kobayashi R., Hirs C.H.W.;
 RT "The amino acid sequence of ovine pancreatic ribonuclease A.";
 RL J. Biol. Chem. 248:7833-7837(1973).
 CC -|- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
 phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
 with 2',3'-cyclic phosphate intermediates.

```

CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S81741; AAB36135.1; -
CC EMBL; S81742; AAB36136.1; -
CC PIR; A00809; NRSH.
CC PIR; A00810; NRGT.
CC HSP; P00656; ISRN.
CC InterPro: IPR001427; RNaseA.
CC Pfam; PF00074; RNaseA; 1.
CC PRINTS; P00794; RIBONUCLEASE.
CC PRODOM; PD000535; RNaseA; 1.
CC SMART; SM00092; RNase_Pc; 1.
CC PROSITE; PS00127; RNASE_PANCREATIC; 1.
CC Hydrolase; Nuclease; Endonuclease; Glycoprotein.
CC FT DISULFID 26
CC FT DISULFID 40
CC FT DISULFID 58
CC FT DISULFID 65
CC FT ACT_SITE 12
CC FT ACT_SITE 12
CC FT ACT_SITE 41
CC FT ACT_SITE 119
CC FT CARBOHYD 34
CC FT CONFLICT 49
CC FT CONFLICT 103
CC FT CONFLICT 103
CC SEQUENCE 124 AA; 13707 MW; 3570EC1452BEDDFA CRC64;
CC
CC Query Match 2.88; Score 12; DB 1; Length 124;
CC Best Local Similarity 100.0%; Pred. No. 0.00012;
CC Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC OY 22 AA AKFERQHMD 33
CC DB 4 AA AKFERQHMD 15
CC
CC RESULT 16
CC RNS_BOVIN STANDARD; PRT; 150 AA.
CC AC P00669;
CC DT 21-JUL-1986 (Rel. 01, Created)
CC DT 01-APR-1990 (Rel. 14, Last sequence update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE Ribonuclease, seminal precursor (EC 3.1.27.5) (Seminal RNase)
CC DE (S-RNase) (Ribonuclease BS-1).
CC GN SRN.
CC OS Bos taurus (Bovine).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
CC OC Bovidae; Bovinae; Bos.
CC OX NCBI_TaxID=9913;
CC RN [1]
CC RN SEQUENCE FROM N.A.
CC RX MEDLINE-90192098; PubMed-2315023;
CC RA Preuss K.D., Wagner S., Freudenstein J., Schelt K.H.;
CC RT "Cloning of cDNA encoding the complete precursor for bovine seminal
CC RT ribonuclease."
CC RL Nucleic Acids Res. 18:1057-1057(1990).
CC RN [2]
CC RN SEQUENCE FROM N.A.
CC RA Sasso M.P., Lombardi M., Confalone E., Carsana A., Palmieri M.,
CC RA Furia A.;
CC RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC [3]

```

```

RP SEQUENCE OF 27-150 FROM N.A.
RX MEDLINE-96139017; PubMed-8587129;
RA Confalone E., Beintema J.J., Sasso M.P., Carsana A., Palmieri M.,
RA Vento M.T., Furia A.;
RT "Molecular evolution of genes encoding ribonucleases in ruminant
RT species."
RL J. Mol. Evol. 41:850-858(1995).
RN [4]
RP SEQUENCE OF 27-150.
RA Suzuki H., Greco L., Parente A., Farina B., la Montagna R., Leone E.;
RT "Primary structure of seminal ribonuclease (RNase BS-1).";
RL Acta Vitaminol. Enzymol. 26:213-214(1972).
RN [5]
RP SEQUENCE OF 73-150 FROM N.A.
RX MEDLINE-86030265; PubMed-3840434;
RA Palmieri M., Carsana A., Furia A., Libonati M.;
RT "Sequence analysis of a cloned cDNA coding for bovine seminal
RT ribonuclease."
RL Eur. J. Biochem. 152:275-277(1985).
RN [6]
RP REVISION TO 43.
RX MEDLINE-83202042; PubMed-6846794;
RA Krietsch W.K.G., Simm F.C., Hertenberger B., Kuntz G.W.K., Wachter E.;
RT "Isolation of bovine seminal ribonuclease by affinity
RT chromatography."
RL Anal. Biochem. 128:213-216(1983).
RN [7]
RP INTERCHAIN DISULFIDE BONDS.
RX MEDLINE-74049879; PubMed-4761089;
RA di Donato A., D'Alessio G.;
RT "Interchain disulfide bridges in ribonuclease BS-1."
RL Biochem. Biophys. Res. Commun. 55:919-928(1973).
RN [8]
RP INTRACHAIN DISULFIDE BONDS.
RX MEDLINE-80130655; PubMed-534646;
RA di Donato A., D'Alessio G.;
RT "Intrachain disulfide bridges of bovine seminal ribonuclease."
RL Biochim. Biophys. Acta 579:303-313(1979).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE-84179398; PubMed-6673761;
RA Capasso S., Giordano F., Mattia C.A., Mazzarella L., Zagari A.;
RT "Refinement of the structure of bovine seminal ribonuclease."
RL Biopolymers 22:327-332(1983).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RA Mazzarella L., Capasso S., Demasi D., di Lorenzo G., Mattia C.A.,
RA Zagari A.;
RT "Bovine seminal ribonuclease: structure at 1.9-A resolution."
RL Acta Crystallogr. D 49:389-402(1993).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (2.06 ANGSTROMS).
RX MEDLINE-99180403; PubMed-10082366;
RA Vitaliano L., Adinolfi S., Riccio A., Sica F., Zagari A.,
RA Mazzarella L.;
RT "Binding of a substrate analog to a domain swapping protein: X-ray
RT structure of the complex of bovine seminal ribonuclease with
RT uridylyl(2',5')adenosine."
RL Protein Sci. 7:1691-1699(1998).
RN [12]
RP MULTIPLE FORMS.
RX MEDLINE-82091855; PubMed-7317378;
RA di Donato A., D'Alessio G.;
RT "Heterogeneity of bovine seminal ribonuclease."
RL Biochemistry 20:7232-7237(1981).
RN [13]
RP ENZYMATIC ACTIVITY.
RX MEDLINE-73213326; PubMed-4664228;
RA D'Alessio G., Parente A., Guida C., Leone E.;
RT "Dimeric structure of seminal ribonuclease."
RL FEBS Lett. 27:285-288(1972).
RN [14]
RP REVIEW.

```

RX MEDLINE-91280661; PubMed-2057997;
RA D'Alessio G., di Donato A., Parente A., Piccoli R.;
RT "Seminal RNase: a unique member of the ribonuclease superfamily.";
RL Trends Biochem. Sci. 16:104-106(1991).
CC -1- FUNCTION: THIS ENZYME HYDROLYZES BOTH SINGLE- AND DOUBLE-STRANDED
CC RNA.
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -1- ENZYME REGULATION: ALLOSTERIC REGULATION BY BOTH SUBSTRATE AND
CC REACTION PRODUCTS.
CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
CC -1- TISSUE SPECIFICITY: SEMINAL PLASMA. CAN REACH 3% OF THE PROTEIN
CC CONTENT OF THIS FLUID.
CC -1- MISCELLANEOUS: PROGRESSIVE DEAMINATION OF ASN-93 TRANSFORMS THE
CC HOMODIMER (BETA- 2) INTO AND HETERODIMER (ALPHA-BETA) AND FINALLY
CC A DOUBLY DEAMINATED DIMER (ALPHA-2).
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X51337; CAA35716.1; -;
CC EMBL; AJ000518; CAA04155.1; -;
CC EMBL; S81747; AAB36140.1; -;
CC EMBL; X03029; CAA26832.1; -;
CC PIR; S08392; NRBO5.
CC PDB; 1BSR; 31-OCT-93.
CC PDB; 11BA; 26-MAR-99.
CC PDB; 11BG; 05-NOV-99.
CC PDB; 1CQJ; 07-MAY-99.
CC PDB; 1CGR; 07-MAY-99.
CC InterPro: IPR001427; RNaseA.
CC Pfam: PF00074; rnasea; 1.
CC PRINTS; PR00794; RIBONUCLEASE.
CC ProDom: PD000535; RNaseA; 1.
CC SMART; SM00092; RNase_Pc; 1.
CC PROSITE; PS00127; RNASE_PANCREATIC; 1.
CC Hydrolase; Nuclease; Endonuclease; Signal; Allosteric enzyme;
CC 3D-structure.
CC
CC SIGNAL 1 26
CC CHAIN 27 150 RIBONUCLEASE, SEMINAL.
CC ACT_SITE 38 38 BY SIMILARITY.
CC ACT_SITE 67 67 BY SIMILARITY.
CC ACT_SITE 145 145 BY SIMILARITY.
CC DISULFID 52 110
CC DISULFID 66 121
CC DISULFID 84 136
CC DISULFID 91 98
CC DISULFID 57 57 INTERCHAIN:
CC DISULFID 58 58 INTERCHAIN.
CC MOD_RES 93 93 DEAMINATION (PROGRESSIVE).
CC HELIX 30 38
CC STRAND 39 39
CC TURN 45 46
CC TURN 48 58
CC TURN 59 60
CC STRAND 69 73
CC HELIX 77 85
CC STRAND 87 89
CC TURN 92 93
CC STRAND 98 100
CC STRAND 105 112
CC TURN 114 115
CC STRAND 117 117
CC TURN 118 119
CC STRAND 120 120
CC STRAND 123 130

FT STRAND 132 137
FT TURN 138 141
FT STRAND 142 149
SQ SEQUENCE 150 AA; 16377 MW; F7A05C930FB83A83 CRC64;

Query Match 2.8%; Score 12; DB 1; Length 150;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAKFERQHMD 33
|||||
DB 30 AAKFERQHMD 41

RESULT 17
RNP_ALCAA STANDARD; PRT; 124 AA.
ID RNP_ALCAA
AC P00667;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE: Ribonuclease pancreatic (EC 3.1.27.5) (RNase A).
GN RNASEL OR RNSI.
OS Alces alces alces (European moose) (Elk).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervioidea;
OC Cervidae; Odocoileinae; Alces.
OX NCBI_TaxID=9853;
RN [1]
RP PARTIAL SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=76003215; PubMed=1157925;
RA Leijenaar-Van den Berg G., Beintema J.J.;
RT "The amino acid sequences of reindeer, moose and fallow deer
RT pancreatic ribonucleases";
RL FEBS Lett. 56:101-107(1975).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC PIR; C91418; NREKN.
CC HSSP; P00636; LSRN.
CC InterPro: IPR001427; RNaseA.
CC Pfam; PF00074; rnasea; 1.
CC PRINTS; PR00794; RIBONUCLEASE.
CC ProDom; PD000535; RNaseA; 1.
CC SMART; SM00092; RNase_Pc; 1.
CC PROSITE; PS00127; RNASE_PANCREATIC; 1.
CC Hydrolase; Nuclease; Endonuclease; Glycoprotein.
CC DISULFID 26 84 BY SIMILARITY.
CC DISULFID 40 95 BY SIMILARITY.
CC DISULFID 58 110 BY SIMILARITY.
CC DISULFID 65 72 BY SIMILARITY.
CC ACT_SITE 12 12 BY SIMILARITY.
CC ACT_SITE 41 41 BY SIMILARITY.
CC ACT_SITE 119 119 BY SIMILARITY.
CC CARBOHYD 34 34 N-LINKED (GLCNAC...).
CC SEQUENCE 124 AA; 13774 MW; 49F9C27F97DE03A7 CRC64;

Query Match 2.6%; Score 11; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAKFERQHMD 32
|||||
DB 4 AAKFERQHMD 14

RESULT 18
RNP_AXIPR STANDARD; PRT; 124 AA.
ID RNP_AXIPR

AC P87351;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Axis porcinus (Hog deer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidae;
OC Cervidae; Cervinae; Axis.
OX NCBI_TaxID=57737;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98278842; PubMed=9611269;
RA Breukelman H.J., van der Munnik N., Kleineidam R.G., Furia A.,
RA Beintema J.J.;
RT "Secretory ribonuclease genes and pseudogenes in true ruminants.";
RL Gene 212:259-268(1998).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y11669; CAA72367.1; -.
DR HSP: P00656; 2RNS.
DR InterPro: IPR001427; RNaseA.
DR Pfam; PF00074; RNaseA; 1.
DR PRINTS; PR00794; RIBONUCLEASE.
DR PRODOM; PD000535; RNaseA; 1.
DR SMART; SM00092; RNase_Pc; 1.
DR PROSITE; PS00127; RNASE_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
SQ SEQUENCE 124 AA; 13750 MW; B182A487553D3915 CRC64;

Query Match 2.6%; Score 11; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AA AKAFERQHMD 32
| | | | | | | | | |
DB 4 AA AKAFERQHMD 14

RESULT 19
RNP_CAPCA STANDARD; PRT; 124 AA.
AC P00664;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Capreolus capreolus (Roe deer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidae;
OC Cervidae; Odocoileinae; Capreolus.

OX NCBI_TaxID=9858;
RN [1]
RP SEQUENCE.
RC TISSUP=Pancreas;
RX MEDLINE=73253236; PubMed=4738402;
RA Zwiers H., Scheffer A.J., Beintema J.J.;
RT "Amino-acid sequences of red-deer and roe-deer pancreatic
RT ribonucleases.";
RL Eur. J. Biochem. 36:569-574(1973).
RN [2]
RP REVISIONS.
RX MEDLINE=77112628; PubMed=836889;
RA Oosterhuis S., Welling G.W., Gaastra W., Beintema J.J.;
RT "Reinvestigation of the primary structures of red deer and roe deer
RT pancreatic ribonuclease and proline sites in mammalian
RT ribonucleases.";
RL Biochim. Biophys. Acta 490:523-529(1977).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=98278842; PubMed=9611269;
RA Breukelman H.J., van der Munnik N., Kleineidam R.G., Furia A.,
RA Beintema J.J.;
RT "Secretory ribonuclease genes and pseudogenes in true ruminants.";
RL Gene 212:259-268(1998).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y11672; CAA72370.1; -.
DR PTR; A90613; NRDEO.
DR HSP: P00656; 1SRN.
DR InterPro: IPR001427; RNaseA.
DR Pfam; PF00074; RNaseA; 1.
DR PRINTS; PR00794; RIBONUCLEASE.
DR PRODOM; PD000535; RNaseA; 1.
DR SMART; SM00092; RNase_Pc; 1.
DR PROSITE; PS00127; RNASE_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease; Glycoprotein.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
FT CARBOHYD 34 34 N-LINKED (GLCNAC...) (PARTIAL).
SQ SEQUENCE 124 AA; 13813 MW; 3836B2017625C613 CRC64;

Query Match 2.6%; Score 11; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AA AKAFERQHMD 32
| | | | | | | | | |
DB 4 AA AKAFERQHMD 14

RESULT 20
RNP_CEREL STANDARD; PRT; 124 AA.
ID RNP_CEREL
AC P00663;
DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase A).
GN RNASE1 OR RNS1.
OS Cervus elaphus (Red deer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidae;
OC Cervidae; Cervinae; Cervus.
OX NCBI_TaxID=9860;
RN [1]
RP SEQUENCE.
RC TISSUE=Pancreas; PubMed=4738402;
RX MEDLINE=7323336; PubMed=4738402;
RA Zwiers H., Scheffer A.J., Beintema J.J.;
RT "Amino-acid sequences of red-deer and roe-deer pancreatic
ribonucleases";
RL Eur. J. Biochem. 36:569-574(1973).
RN [2]
RP REVISIONS.
RX MEDLINE=77112528; PubMed=836889;
RA Oosterhuis S., Wellings G.W., Gastra W., Beintema J.J.;
RT "Reinvestigation of the primary structures of red deer and roe deer
pancreatic ribonuclease and proline sites in mammalian
ribonucleases";
RL Biochim. Biophys. Acta 490:523-529(1977).
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
with 2',3'-cyclic phosphate intermediates.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: PANCREAS.
CC -!- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
DR PIR; B90613; NRDEF.
DR HSSP; P00656; ISRN.
DR InterPro; IPR001427; RNaseA.
DR Pfam; PF00074; rnaA; 1.
DR PRINTS; PR00794; RIBONUCLEASE.
DR ProDom; PD000535; RNaseA; 1.
DR SMART; SM00092; RNase_PC; 1.
DR PROSITE; PS00127; RNase_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
SQ SEQUENCE 124 AA; 13777 MW; 330F86D52B3DD71B CRC64;

Query Match 2.6%; Score 11; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. NO. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAKFERQHMD 32
DB 4 AAKFERQHMD 14

RESULT 21
RNP_DAMDA
ID RNP_DAMDA STANDARD; PRT; 124 AA.
AC P00665;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase A).
GN RNASE1 OR RNS1.
OS Dama dama (fallow deer) (Cervus dama).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidae;
OC Cervidae; Cervinae; Cervus.
OX NCBI_TaxID=30532;
RN [1]

RP PARTIAL SEQUENCE.
RX MEDLINE=76003215; PubMed=1157925;
RA Leijenaar-Van den Berg G., Beintema J.J.;
RT "The amino acid sequences of reindeer, moose and fallow deer
pancreatic ribonucleases";
RL FEBS Lett. 56:101-107(1975).
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
with 2',3'-cyclic phosphate intermediates.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: PANCREAS.
CC -!- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
DR PIR; A91418; NRDEF.
DR HSSP; P00656; ISRN.
DR InterPro; IPR001427; RNaseA.
DR Pfam; PF00074; rnaA; 1.
DR PRINTS; PR00794; RIBONUCLEASE.
DR ProDom; PD000535; RNaseA; 1.
DR SMART; SM00092; RNase_PC; 1.
DR PROSITE; PS00127; RNase_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
SQ SEQUENCE 124 AA; 13807 MW; A871BE3BFACE1827 CRC64;

Query Match 2.6%; Score 11; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. NO. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAKFERQHMD 32
DB 4 AAKFERQHMD 14

RESULT 22
RNP_RANTA
ID RNP_RANTA STANDARD; PRT; 124 AA.
AC P00666;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase A).
GN RNASE1 OR RNS1.
OS Rangifer tarandus (Reindeer) (Caribou).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidae;
OC Cervidae; Odocoileinae; Rangifer.
OX NCBI_TaxID=9870;
RN [1]
RP PARTIAL SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=76003215; PubMed=1157925;
RA Leijenaar-Van den Berg G., Beintema J.J.;
RT "The amino acid sequences of reindeer, moose and fallow deer
pancreatic ribonucleases";
RL FEBS Lett. 56:101-107(1975).
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
with 2',3'-cyclic phosphate intermediates.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: PANCREAS.
CC -!- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
DR PIR; B91418; NRDEF.
DR HSSP; P00656; ISRN.
DR InterPro; IPR001427; RNaseA.
DR Pfam; PF00074; rnaA; 1.
DR PRINTS; PR00794; RIBONUCLEASE.
DR ProDom; PD000535; RNaseA; 1.

```

DR SMART; SM00092; RNase_Pc; 1.
DR PROSITE; PS00127; RNASE_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
FT VARIANT 103 103
SQ SEQUENCE 124 AA; 13759 MW; C5568B8B9A87CD54 CRC64;

Query Match 2.6%; Score 11; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 AA:KFERQHMD 32
DB 4 AA:KFERQHMD 14

RESULT 23
RNP_CAMDR
ID RNP_CAMDR STANDARD; PRT; 124 AA;
AC P00670;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Camelus dromedarius (Dromedary) (Arabian camel), and
OS Camelus bactrianus (Bactrian camel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX NCBI_TaxID=9838, 9837;
RN [1]
RP SEQUENCE.
RC SPECIES=C.dromedarius;
RX MEDLINE=76039472; PubMed=1167157;
RA Welling G.W., Groen G., Beintema J.J.;
RT "The amino acid sequence of dromedary pancreatic ribonuclease.";
RL Biochem. J. 147:505-511(1975).
RN [2]
RP MAJOR VARIANT, PARTIAL SEQUENCE, AND REVISIONS.
RC SPECIES=C.dromedarius;
RX MEDLINE=85204378; PubMed=3922790;
RA Beintema J.J.;
RT "Mammalian ribonucleases. The absence of a glycosylated Asn-Pro-Thr
RT sequence in horse ribonuclease and the presence of tryptophan at
RT position 39 in horse and dromedary ribonuclease.";
RL FEBS Lett. 185:115-120(1985).
RN [3]
RP PARTIAL SEQUENCE, AND VARIANT ALLELIC.
RC SPECIES=C.bactrianus;
RX MEDLINE=76277807; PubMed=962846;
RA Welling G.W., Mulder H., Beintema J.J.;
RT "Allelic polymorphism in arabian camel ribonuclease and the amino
RT acid sequence of bactrian camel ribonuclease.";
RL Biochem. Genet. 14:309-317(1976).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
DR PIR: A00815; NRCM.
DR PIR: A90229; NRCMM.
DR PIR: B90229; NRCMB.
DR HSSP: P00656; ISRN.
DR InterPro: IPR001427; RNaseA.
DR Pfam: PF00074; rnasea; 1.
DR PRINTS: PR00794; RIBONUCLEASE.
DR ProDom: PD000535; RNaseA; 1.

```

```

DR SMART; SM00092; RNase_Pc; 1.
DR PROSITE; PS00127; RNASE_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
FT VARIANT 103 103
SQ SEQUENCE 124 AA; 13938 MW; 503ABA0BAAEFE92F CRC64;

Query Match 2.1%; Score 9; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 KFERQHMD 33
DB 7 KFERQHMD 15

RESULT 24
RNP_GIRCA
ID RNP_GIRCA STANDARD; PRT; 124 AA.
AC P00662; Q29541; Q29534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Giraffa camelopardalis (Giraffe).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Giraffoidea;
OC Giraffidae; Giraffa.
OX NCBI_TaxID=9894;
RN [1]
RP SEQUENCE.
RA Gastra W.;
RL Thesis (1975), University of Groningen, Netherland.
RN [2]
RP SEQUENCE.
RX MEDLINE=74309061; PubMed=4855009;
RA Gastra W., Groen G., Welling G.W., Beintema J.J.;
RT "The primary structure of giraffe pancreatic ribonuclease.";
RL FEBS Lett. 41:227-232(1974).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96139017; PubMed=8587129;
RA Confalone E., Beintema J.J., Sasso M.P., Carsana A., Palmieri M.,
RA Vento M.T., Furia A.;
RT "Molecular evolution of genes encoding ribonucleases in ruminant
RT species.";
RL J. Mol. Evol. 41:850-858(1995).
RN [4]
RP SEQUENCE OF 31-114 FROM N.A.
RX MEDLINE=93367815; PubMed=8360916;
RA Breukelman H.J., Beintema J.J., Confalone E., Costanzo C., Sasso M.P.,
RA Carsana A., Palmieri M., Furia A.;
RT "Sequences related to the ox pancreatic ribonuclease coding region in
RT the genomic DNA of mammalian species.";
RL J. Mol. Evol. 37:29-35(1993).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```


use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

CC EMBL; S81739; AAB36133.1; -;
CC EMBL; S65127; AAB27932.1; -;
CC PIR; A00811; NRGF.
CC HSSP; P00656; ISSB.
CC InterPro; IPR001427; RNaseA.
CC Pfam; PF00074; rnaaseA; 1.
CC PRINTS; PR00794; RIBONUCLEASE.
CC ProDom; PD000535; RNaseA; 1.
CC SMART; SM00092; RNase_PC; 1.
CC PROSITE; PS00127; RNASE_PANCREATIC; 1.
CC Hydrolase; Nuclease; Endonuclease; Glycoprotein.
KW DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .).
FT VARIANT 76 76 Y -> N.
FT CONFLICT 28 28 E -> Q (IN REF. 1 AND 2).
SQ SEQUENCE 124 AA; 13704 MW; CE97EBDC792612DA CRC64;

Query Match 2.1%; Score 9; DB 1; Length 124;

Best Local Similarity 100.0%; Pred. No. 0.14;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AAKEFERQH 30

DB 4 AAKEFERQH 12

RESULT 25

RNP_MESAU

ID RNP_MESAU STANDARD; PRT; 124 AA.

AC P00682;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).

GN RNASE1 OR RNS1.

OS Mesocricetus auratus (Golden hamster).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;

OC Mesocricetus.

OX NCBI_TaxID=10036;

RN [1]

RP SEQUENCE.

RC TISSUE=Pancreas;

RA MEDLINE=80088445; PubMed=518928;

RA Jekel P.A., Lips H.J., Lenstra J.A., Beintema J.J.;

RT "The amino acid sequence of hamster pancreatic ribonuclease.";

RL Biochimie 61:827-839(1979).

CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-

CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P

CC with 2',3'-cyclic phosphate intermediates.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- TISSUE SPECIFICITY: PANCREAS.

CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.

DR PIR; A00829; NRHY.

DR HSSP; P00656; ISRN.

DR InterPro; IPR001427; RNaseA.

DR Pfam; PF00074; rnaaseA; 1.

DR PRINTS; PR00794; RIBONUCLEASE.

DR ProDom; PD000535; RNaseA; 1.

DR SMART; SM00092; RNase_PC; 1.

DR PROSITE; PS00127; RNASE_PANCREATIC; 1.

KW Hydrolase; Nuclease; Endonuclease; Glycoprotein.

FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .).
SQ SEQUENCE 124 AA; 13867 MW; 4E3C57F94E73C478 CRC64;

Query Match

Best Local Similarity 100.0%; Pred. No. 0.14;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33

DB 7 KFERQHMS 15

RESULT 26

RNP_ONDZI

ID RNP_ONDZI STANDARD; PRT; 124 AA.

AC P00681;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).

GN RNASE1 OR RNS1.

OS Ondatra zibethicus (Muskrat).

OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae;

OC Ondatra.

OX NCBI_TaxID=10060;

RN [1]

RP SEQUENCE.

RC TISSUE=Pancreas;

RA MEDLINE=76212474; PubMed=1279085;

RA van Dijk H., Sloots B., van den Berg A., Gaastria W., Beintema J.J.;

RT "The primary structure of muskrat pancreatic ribonuclease.";

RL Int. J. Pept. Protein Res. 8:305-316(1976).

CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-

CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P

CC with 2',3'-cyclic phosphate intermediates.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- TISSUE SPECIFICITY: PANCREAS.

CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.

DR PIR; A00828; NROZ.

DR HSSP; P00656; ISRN.

DR InterPro; IPR001427; RNaseA.

DR Pfam; PF00074; rnaaseA; 1.

DR PRINTS; PR00794; RIBONUCLEASE.

DR ProDom; PD000535; RNaseA; 1.

DR SMART; SM00092; RNase_PC; 1.

DR PROSITE; PS00127; RNASE_PANCREATIC; 1.

KW Hydrolase; Nuclease; Endonuclease.

FT DISULFID 26 84 BY SIMILARITY.

FT DISULFID 40 95 BY SIMILARITY.

FT DISULFID 58 110 BY SIMILARITY.

FT DISULFID 65 72 BY SIMILARITY.

FT ACT_SITE 12 12 BY SIMILARITY.

FT ACT_SITE 41 41 BY SIMILARITY.

FT ACT_SITE 119 119 BY SIMILARITY.

FT UNSURE 32 33

FT UNSURE 75 76

SQ SEQUENCE 124 AA; 13880 MW; A9E42DC90E9E7C04 CRC64;

Query Match

Best Local Similarity 100.0%; Pred. No. 0.14;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33

DB 7 KFERQHMS 15

```

RESULT 27
RNP_HORSE
ID RNP_HORSE STANDARD; PRT; 128 AA.
AC P00674;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN PRELIMINARY SEQUENCE.
RP MEDLINE=74302367; PubMed=4852291;
RX Scheffer A.J., Beintema J.J.;
RT "Horse pancreatic ribonuclease.";
RL Eur. J. Biochem. 46:221-233(1974).
[2]
RP REVISIONS.
RX MEDLINE=85204378; PubMed=3922790;
RA Beintema J.J.;
RT "Mammalian ribonucleases. The absence of a glycosylated Asn-Pro-Thr
sequence in horse ribonuclease and the presence of tryptophan at
position 39 in horse and dromedary ribonuclease.";
RL FEBS Lett. 185:115-120(1985).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
PIR: A00819; NRHO.
DR HSP; P00656; ISRN.
DR InterPro: IPR001427; RNaseA.
DR PRINTS; PR00794; RNaseA; 1.
DR PRODOM; PD000535; RNaseA; 1.
DR SMART; SM00092; RNase_Pc; 1.
DR PROSITE; PS00127; RNASE_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease; Glycoprotein.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
FT CARBOHYD 34 34 N-LINKED (GLCNAC...).
FT VARIANT 98 98 G -> R.
FT SEQUENCE 128 AA; 14374 MW; A06727414097CDD CRC64;
SQ SEQUENCE 128 AA; 14273 MW; A213AE40786DF9AE CRC64;

Query Match 2.1%; Score 9; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33
Db 7 KFERQHMS 15

RESULT 29
RNP_MYOCO
ID RNP_MYOCO STANDARD; PRT; 128 AA.
AC P00676;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Myocastor coypus (Coypu) (Nutria).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Myocastoridae;
OC Myocastor.
OX NCBI_TaxID=10157;
RN PRELIMINARY SEQUENCE.
RP MEDLINE=77065676; PubMed=999896;
RX van den Berg A., van den Hende-Timmer L., Beintema J.J.;
RT "Isolation, properties and primary structure of coypu and chinchilla
pancreatic ribonuclease.";
RL Biochim. Biophys. Acta 453:400-409(1976).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.

Query Match 2.1%; Score 9; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33
Db 7 KFERQHMS 15

RESULT 28
RNP_HYSCR
ID RNP_HYSCR STANDARD; PRT; 128 AA.
AC P04060;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-NOV-1986 (Rel. 03, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.

```

```

OS Hystrix cristata (Crested porcupine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Hystricidae; Hystrix.
OX NCBI_TaxID=10137;
RN PRELIMINARY SEQUENCE.
RP MEDLINE=83000399; PubMed=7115727;
RX Beintema J.J., Knol G., Martena B.;
RT "The primary structures of pancreatic ribonucleases from African
porcupine and casiragua, two hystricomorph rodent species.";
RL Biochim. Biophys. Acta 705:102-110(1982).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
PIR: A00823; NRPO.
DR HSP; P00656; ISRN.
DR InterPro: IPR001427; RNaseA.
DR PRINTS; PR00794; RNaseA; 1.
DR PRODOM; PD000535; RNaseA; 1.
DR SMART; SM00092; RNase_Pc; 1.
DR PROSITE; PS00127; RNASE_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease; Glycoprotein.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
FT CARBOHYD 34 34 N-LINKED (GLCNAC...).
FT VARIANT 98 98 G -> R.
FT SEQUENCE 128 AA; 14273 MW; A213AE40786DF9AE CRC64;

Query Match 2.1%; Score 9; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33
Db 7 KFERQHMS 15

RESULT 29
RNP_MYOCO
ID RNP_MYOCO STANDARD; PRT; 128 AA.
AC P00676;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Myocastor coypus (Coypu) (Nutria).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Myocastoridae;
OC Myocastor.
OX NCBI_TaxID=10157;
RN PRELIMINARY SEQUENCE.
RP MEDLINE=77065676; PubMed=999896;
RX van den Berg A., van den Hende-Timmer L., Beintema J.J.;
RT "Isolation, properties and primary structure of coypu and chinchilla
pancreatic ribonuclease.";
RL Biochim. Biophys. Acta 453:400-409(1976).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.

```

CC -!- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
DR PIR: A00822; NRCU.
DR HSP: P00656; 1SRN.
DR InterPro: IPR001427; RNaseA.
DR Pfam: PF00074; rnasea; 1.
DR PRINTS: PR00794; RIBONUCLEASE.
DR PRODOM: PD000535; RNaseA; 1.
DR SMART: SM00092; RNase_Pc; 1.
DR KW Hydrolyase; Nuclease; Endonuclease; Glycoprotein.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
FT CARBOHYD 34 34 N-LINKED (GLCNAC...).
SQ SEQUENCE 128 AA; 14267 MW; 4EB924E5B445832 CRC64;

Query Match 2.1%; Score 9; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33
|||||
DB 7 KFERQHMS 15

RESULT 30

ID RNP_CRILLO STANDARD; PRT; 130 AA.
AC P24717;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic precursor (EC 3.1.27.5) (RNase 1) (RNase A) (Fragment).
GN RNASE1 OR NS1.
OS Cricetus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10030;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RX MEDLINE=92158677; PubMed=1741299;
RA Haug M., Schein C.H.;
RT "The DNA sequences of the human and hamster secretory ribonucleases determined with the polymerase chain reaction (PCR).";
RL Nucleic Acids Res. 20:612-612(1992).
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-phosphates and 3'-phosphooligonucleotides ending in C-P or U-P with 2',3'-cyclic phosphate intermediates.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: PANCREAS.
CC -!- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62945; CAA44717.1;
DR PIR: S22808; S22808.
DR HSP: P00656; 2RNS.
DR InterPro: IPR001427; RNaseA.
DR Pfam: PF00074; rnasea; 1.
DR PRODOM: PD000535; RNaseA; 1.

DR SMART; SM00092; RNase_Pc; 1.
DR PROSITE; PS00127; RNASE_PANCREATIC; 1.
KW Hydrolyase; Nuclease; Endonuclease; Glycoprotein; Signal.
FT NON_TER 1 1
FT SIGNAL <1 6
FT CHAIN 7 130 RIBONUCLEASE PANCREATIC.
FT DISULFID 32 90 BY SIMILARITY.
FT DISULFID 46 101 BY SIMILARITY.
FT DISULFID 64 116 BY SIMILARITY.
FT DISULFID 71 78 BY SIMILARITY.
FT ACT_SITE 18 18 BY SIMILARITY.
FT ACT_SITE 47 47 BY SIMILARITY.
FT ACT_SITE 125 125 BY SIMILARITY.
FT CARBOHYD 40 40 N-LINKED (GLCNAC...). (PROBABLE).
SQ SEQUENCE 130 AA; 14517 MW; 8D227FD6491B413D CRC64;

Query Match 2.1%; Score 9; DB 1; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33
|||||
DB 13 KFERQHMS 21

RESULT 31

ID RNP_CLEGL STANDARD; PRT; 148 AA.
AC Q9WUR4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic precursor (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1.
OS Clethrionomys glareolus (Bank vole).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae;
OC Clethrionomys.
OX NCBI_TaxID=51090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99439681; PubMed=10508551;
RA Dubois J.-Y.F., Catzeffis F.M., Beintema J.J.;
RT "The phylogenetic position of 'Acomyinae' (Rodentia, Mammalia) as sister group of a Murinae + Gerbillinae clade: evidence from the nuclear ribonuclease gene.";
RL Mol. Phylogenet. Evol. 13:181-192(1999).
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-phosphates and 3'-phosphooligonucleotides ending in C-P or U-P with 2',3'-cyclic phosphate intermediates.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: PANCREAS.
CC -!- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ005769; CAB41470.1;
DR HSP: P00656; 1SRN.
DR InterPro: IPR001427; RNaseA.
DR Pfam: PF00074; rnasea; 1.
DR PRINTS: PR00794; RIBONUCLEASE.
DR PRODOM: PD000535; RNaseA; 1.
DR SMART; SM00092; RNase_Pc; 1.
DR PROSITE; PS00127; RNASE_PANCREATIC; 1.
KW Hydrolyase; Nuclease; Endonuclease; Glycoprotein; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 148 RIBONUCLEASE PANCREATIC.

```
FT DISULFID 50 108 BY SIMILARITY..
FT DISULFID 64 119 BY SIMILARITY.
FT DISULFID 82 134 BY SIMILARITY.
FT DISULFID 89 96 BY SIMILARITY.
FT ACT_SITE 37 37 BY SIMILARITY.
FT ACT_SITE 65 65 BY SIMILARITY.
FT ACT_SITE 143 143 BY SIMILARITY.
FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 148 AA; 16471 MW; 028BD249C2DDB563 CRC64;

Query Match 2.1%; Score 9; DB 1; Length 148;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33
DB 32 KFERQHMS 40
|||||

RESULT 32
RNP_GERNI STANDARD; PRT; 148 AA.
AC Q9WU52;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic precursor (EC 3.1.27.5) (RNase A).
GN RNASE1.
OS Gerbillus nigeriae (Nigerian gerbil).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;
OC Gerbillus.
OX NCBI_TaxID=39472;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99439681; PubMed=10508551;
RA Dubois J.-Y.F., Catzeffis F.M., Beintema J.J.;
RT "The phylogenetic position of 'Acomyinae' (Rodentia, Mammalia) as
RT sister group of a Murinae + Gerbillinae clade: evidence from the
RT nuclear ribonuclease gene.";
RL Mol. Phylogenet. Evol. 13:181-192(1999).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ005774; CAB41479.1;
CC HSSP; P00656; ISRN.
CC InterPro; IPR001427; RNaseA.
CC Pfam; PF00074; rnasea; 1.
CC PRINTS; PR00794; RIBONUCLEASE.
CC ProDom; PD000535; RNaseA; 1.
CC SMART; SM00092; RNase_Pc; 1.
CC PROSITE; PS00127; RNASE_PANCREATIC; 1.
CC KEGG; K00002; RNase; Signal.
CC SIGNAL 1 25 POTENTIAL.
CC CHAIN 26 148 RIBONUCLEASE PANCREATIC.
CC DISULFID 50 108 BY SIMILARITY.
CC DISULFID 64 119 BY SIMILARITY.
CC DISULFID 82 134 BY SIMILARITY.
CC DISULFID 89 96 BY SIMILARITY.
CC ACT_SITE 37 37 BY SIMILARITY.
CC ACT_SITE 65 65 BY SIMILARITY.
CC ACT_SITE 143 143 BY SIMILARITY.
SQ SEQUENCE 148 AA; 16331 MW; 264FE2DDEC351F9F CRC64;

Query Match 2.1%; Score 9; DB 1; Length 148;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
FT ACT_SITE 65 65 BY SIMILARITY.
FT ACT_SITE 143 143 BY SIMILARITY.
SQ SEQUENCE 148 AA; 16348 MW; 154EFF5AE58C6B9A CRC64;

Query Match 2.1%; Score 9; DB 1; Length 148;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33
DB 32 KFERQHMS 40
|||||

RESULT 33
RNP_TATKG STANDARD; PRT; 148 AA.
AC Q9WUX4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic precursor (EC 3.1.27.5) (RNase A).
GN RNASE1.
OS Tatera kempii gambiana (Kemp's gerbil).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;
OC Tatera.
OX NCBI_TaxID=41264;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99439681; PubMed=10508551;
RA Dubois J.-Y.F., Catzeffis F.M., Beintema J.J.;
RT "The phylogenetic position of 'Acomyinae' (Rodentia, Mammalia) as
RT sister group of a Murinae + Gerbillinae clade: evidence from the
RT nuclear ribonuclease gene.";
RL Mol. Phylogenet. Evol. 13:181-192(1999).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ005775; CAB41485.1;
CC HSSP; P00656; ISSA.
CC InterPro; IPR001427; RNaseA.
CC Pfam; PF00074; rnasea; 1.
CC PRINTS; PR00794; RIBONUCLEASE.
CC ProDom; PD000535; RNaseA; 1.
CC SMART; SM00092; RNase_Pc; 1.
CC PROSITE; PS00127; RNASE_PANCREATIC; 1.
CC KEGG; K00002; RNase; Signal.
CC SIGNAL 1 25 POTENTIAL.
CC CHAIN 26 148 RIBONUCLEASE PANCREATIC.
CC DISULFID 50 108 BY SIMILARITY.
CC DISULFID 64 119 BY SIMILARITY.
CC DISULFID 82 134 BY SIMILARITY.
CC DISULFID 89 96 BY SIMILARITY.
CC ACT_SITE 37 37 BY SIMILARITY.
CC ACT_SITE 65 65 BY SIMILARITY.
CC ACT_SITE 143 143 BY SIMILARITY.
SQ SEQUENCE 148 AA; 16331 MW; 264FE2DDEC351F9F CRC64;

Query Match 2.1%; Score 9; DB 1; Length 148;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 25 KFERQHMS 33
DB 32 KFERQHMS 40

RESULT 34
RNP_ACOCA
ID RNP_ACOCA STANDARD; PRT; 149 AA.
AC Q9WTT5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic precursor (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1.
OS Acomys cahirinus (Egyptian spiny mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Acomys.
OX NCBI_TaxID=10068;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99439681; PubMed=10508551;
RA Dubois J.-Y.F., Catzeflis F.M., Beintema J.J.;
RT "The phylogenetic position of 'Acomyinae' (Rodentia, Mammalia) as
RT sister group of a Murinae + Gerbillinae clade: evidence from the
RT nuclear ribonuclease gene.";
RL Mol. Phylogenet. Evol. 13:181-192(1999).
CC -|- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- TISSUE SPECIFICITY: PANCREAS.
CC -|- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ005772; CAB41464.1;
CC HSP: P00656; 1SRN.
CC InterPro: IPR001427; RNaseA.
CC Pfam: PF00074; RNaseA.
CC PRINTS: PR00794; RIBONUCLEASE.
CC ProDom: PD000535; RNaseA.
CC SMART: SM00092; RNase_PC.
CC PROSITE: PS00127; RNase_PC.
CC Hydrolyase; Nuclease; Endonuclease; Glycoprotein; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 149 RIBONUCLEASE PANCREATIC.
FT DISULFID 51 109 BY SIMILARITY.
FT DISULFID 65 120 BY SIMILARITY.
FT DISULFID 83 135 BY SIMILARITY.
FT DISULFID 90 97 BY SIMILARITY.
FT ACT_SITE 37 37 BY SIMILARITY.
FT ACT_SITE 66 66 BY SIMILARITY.
FT ACT_SITE 144 144 BY SIMILARITY.
FT ACT_SITE 144 144 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 87 87 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 149 AA; 16613 MW; 9A25B5C11040456C CRC64;

Query Match 2.1%; Score 9; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33
DB 32 KFERQHMS 40

RESULT 35
RNP_MUSPA
ID RNP_MUSPA STANDARD; PRT; 149 AA.
AC Q9QYX3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Ribonuclease pancreatic precursor (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1.
OS Akodon jelskii.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Sigmodontinae;
OX NCBI_TaxID=10079;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99439681; PubMed=10508551;
RA Dubois J.-Y.F., Catzeflis F.M., Beintema J.J.;
RT "The phylogenetic position of 'Acomyinae' (Rodentia, Mammalia) as
RT sister group of a Murinae + Gerbillinae clade: evidence from the
RT nuclear ribonuclease gene.";
RL Mol. Phylogenet. Evol. 13:181-192(1999).
CC -|- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC with 2',3'-cyclic phosphate intermediates.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- TISSUE SPECIFICITY: PANCREAS.
CC -|- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ005771; CAB41465.1;
CC HSP: P00656; 2RNS.
CC InterPro: IPR001427; RNaseA.
CC Pfam: PF00074; RNaseA.
CC PRINTS: PR00794; RIBONUCLEASE.
CC ProDom: PD000535; RNaseA.
CC SMART: SM00092; RNase_PC.
CC PROSITE: PS00127; RNase_PC.
CC Hydrolyase; Nuclease; Endonuclease; Glycoprotein; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 149 RIBONUCLEASE PANCREATIC.
FT DISULFID 51 109 BY SIMILARITY.
FT DISULFID 65 120 BY SIMILARITY.
FT DISULFID 83 135 BY SIMILARITY.
FT DISULFID 90 97 BY SIMILARITY.
FT ACT_SITE 37 37 BY SIMILARITY.
FT ACT_SITE 66 66 BY SIMILARITY.
FT ACT_SITE 144 144 BY SIMILARITY.
FT ACT_SITE 144 144 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 87 87 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 149 AA; 16611 MW; 20F9CBE9089ABBD3 CRC64;

Query Match 2.1%; Score 9; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 KFERQHMS 33
DB 32 KFERQHMS 40

RESULT 36
RNP_MUSPA
ID RNP_MUSPA STANDARD; PRT; 149 AA.
AC Q9QYX3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
```


RT "Protein Hpn: cloning and characterization of a histidine-rich metal-
binding polypeptide in Helicobacter pylori and Helicobacter
mustelae.";
RL Infect. Immun. 63:2682-2688(1995).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervilange A.R., Clayton R.A., Sutton G.G.,
Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
Berg D.E., Gocayne J.D., Uckerback T.R., Peterson J.D., Kelley J.M.,
Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
Venter J.C.;
RA "The complete genome sequence of the gastric pathogen Helicobacter
pylori.";
RT Nature 388:539-547(1997).
RL [3]

RP SEQUENCE FROM N.A.
RC STRAIN=J99;
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
Smith D.R., Noonan B., Guille B.C., deJonge B.L., Carmel G.,
Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
Trust T.J.;
RA "Genomic sequence comparison of two unrelated isolates of the human
gastric pathogen Helicobacter pylori.";
RT Nature 397:176-180(1999).

RL -1- FUNCTION: STRONGLY BINDS NICKEL AND ZINC. BINDS OTHER METALS LESS
CC STRONGLY: CO2+ > CU2+ > MN2+. MAY ACT TO INCREASE, OR AT
CC LEAST TO PRESERVE, UREASE ACTIVITY. EXACT FUNCTION IS STILL
CC UNKNOWN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U26361; AAA85859.1; -;
DR EMBL; AEO00643; AAD08471.1; -;
DR EMBL; AEO01555; AAD06898.1; -;
DR TIGR; HPI427; -;
KW Metal-binding; Zinc; Nickel; Repeat; Complete proteome.
FT INIT_MET 0 0
FT DOMAIN 10 23 POLY-HIS.
FT DOMAIN 27 32 POLY-HIS.
FT DOMAIN 37 54 2 X 5 AA REPEATS OF E-E-G-C-C.
FT REPEAT 37 41 1.
FT REPEAT 50 54 2.
SQ SEQUENCE 59 AA; 6946 MW; C3AEE3F602EC973C CRC64;

Query Match 1.9%; Score 8; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 0.73;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HHHHHHSS 9
Db 27 HHHHHHSS 34

RESULT 39
ID Vxis_BPL54 STANDARD; PRT; 59 AA.
AC P20710;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 01-MAY-1991 (Rel. 18, Last annotation update)
DE Excisionase.
GN XIS
OS Bacteriophage L54a.
OC Viruses.
OX NCBI_TaxID=12373;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89327124; PubMed=2526804;
RA Ye Z.-H., Lee C.Y.;
RT "Nucleotide sequence and genetic characterization of staphylococcal
bacteriophage L54a int and xis genes.";
RL J. Bacteriol. 171:4146-4153(1989).
CC -1- FUNCTION: EXCISIONASE AND INTEGRASE ARE NECESSARY FOR THE EXCISION
CC OF PROPHAGE FROM THE HOST GENOME BY SITE-SPECIFIC RECOMBINATION AT
CC THE ATT SITE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M27965; AAA98160.1; -;
DR PIR; A33855; RSBPL5.
KW DNA recombination; DNA-binding.
SQ SEQUENCE 59 AA; 7199 MW; B1815CC035AA773E CRC64;

Query Match 1.9%; Score 8; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 0.73;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 YIKESVEN 167
Db 12 YIKESVEN 19

RESULT 40
EOTA_MOUSE

ID EOTA_MOUSE STANDARD; PRT; 97 AA.
AC P48298; -;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Eotaxin precursor (Small inducible cytokine All) (CCL11) (Eosinophil
DE chemotactic protein).
GN SCYALL.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Lung;
RX MEDLINE=96004658; PubMed=7568052;

RA Rothenberg M.E., Luster A.D., Leder P.;

RT "Murine eotaxin: an eosinophil chemoattractant inducible in
RT endothelial cells and in interleukin 4-induced tumor suppression.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:8960-8964(1995).
RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Lung;

RX MEDLINE=96158746; PubMed=8574847;

RA Gonzalo J.-A., Jia G.-O., Aguirre V., Friend D., Coyle A.J.,

RA Jenkins N.A., Lin G.-S., Katz H., Lichtman A., Copeland N.G., Kopf M.,

RA Gutierrez-Ramos J.-C.;

RT "Mouse eotaxin expression parallels eosinophil accumulation during
RT lung allergic inflammation but it is not restricted to a Th2-type
RT response.";

RL Immunity 4:1-14(1996).

CC -1- FUNCTION: IN RESPONSE TO THE PRESENCE OF ALLERGENS, THIS PROTEIN


```

CC DIRECTLY PROMOTES THE ACCUMULATION OF EOSINOPHILS (A PROMINENT
CC FEATURE OF ALLERGIC INFLAMMATORY REACTIONS), BUT NOT
CC LYMPHOCYTES, MACROPHAGES OR NEUTROPHILS.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- TISSUE SPECIFICITY: EXPRESSED CONSTITUTIVELY IN THE THYMUS.
CC EXPRESSION INDUCIBLE IN THE LUNG (TYPE I-ALVEOLAR EPITHELIAL
CC CELLS), INTESTINE, HEART, SPLEEN, KIDNEY.
CC -|- INDUCTION: BY INTERFERON GAMMA AND LIPOPOLYSACCHARIDE (LPS).
CC -|- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
CC C-C) (CHEMOKINE CC).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U26426; AAC52256.1; -;
CC EMBL; U40672; AAA99776.1; -;
CC HSPF; P51671; LEOT.
CC MGD; MGI:103576; Scyall.
CC InterPro; IPR000827; CC_chemkine_sml.
CC InterPro; IPR001811; Chemokine_IL8.
CC Pfam; PF00048; IL8; 1.
CC SMART; SM00199; SCY; 1.
CC PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
KW Eosinophil; Cytokine; Chemotaxis; Glycoprotein; Signal;
KW Inflammatory response.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 97 EOTAXIN.
FT DISULFID 32 57 BY SIMILARITY.
FT DISULFID 33 73 BY SIMILARITY.
FT CARBOHYD 94 94 O-LINKED (GALNAC...) (BY SIMILARITY).
SQ SEQUENCE 97 AA; 10893 MW; 36C9812107FC6CA7 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 152 NTLKSYK 159
Db 43 NTLKSYK 50

```

Search completed: May 12, 2003, 10:26:40
Job time : 42 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 10:23:33 ; Search time 35 Seconds
(without alignments)
2537.323 Million cell updates/sec

Title: US-10-057-531A-2
Perfect score: 431
Sequence: 1 MHHHHHHHSLVPRSGMKE.....TCECTKPDVPLFDGIFCSS 431

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	374	86.8	1720	5	Q25922 plasmodium
2	344	79.8	652	5	Q25923 plasmodium
3	293	68.0	373	5	Q25723 plasmodium
4	231	53.6	373	5	O43995 plasmodium
5	231	53.6	373	5	O43996 plasmodium
6	231	53.6	373	5	Q25724 plasmodium
7	192	44.5	373	5	Q25722 plasmodium
8	190	44.1	570	5	O9TYG2 plasmodium
9	190	44.1	570	5	Q25968 plasmodium
10	189	43.9	373	5	Q25721 plasmodium
11	165	38.3	373	5	Q25727 plasmodium
12	165	38.3	373	5	Q25728 plasmodium
13	145	33.6	372	5	O43997 plasmodium
14	145	33.6	372	5	Q25725 plasmodium
15	145	33.6	569	5	Q25978 plasmodium
16	145	33.6	569	5	Q25983 plasmodium

90	15	3.5	134	6	Q9T5F2	Q9tsf2 bos taurus	163	7	1.6	134	8	Q9GDC5	Q9gdg5 malus loens
91	15	3.5	674	5	Q9GRE8	Q9gre8 toxoplasma	164	7	1.6	134	8	Q9GDF7	Q9gdg2 malus yunna
92	12	2.8	21	6	Q9TRQ8	Q9trq8 bos taurus	165	7	1.6	135	8	Q9GDF2	Q9gdg7 malus fusca
93	12	2.8	124	6	Q9M235	Q9m235 bubalus bub	166	7	1.6	136	4	Q9NXL9	Q9nxl9 homo sapien
94	12	2.8	144	6	Q9BH14	Q9bh14 antilocapra	167	7	1.6	136	4	Q9HCV5	Q9hcv5 homo sapien
95	9	2.1	116	6	Q9TVB7	Q9tvb7 camelus dro	168	7	1.6	136	8	Q9GIF1	Q9glf1 malus sieve
96	9	2.1	134	6	Q9BDB9	Q9bdb9 tragulus ja	169	7	1.6	136	8	Q9GIF0	Q9glf0 malus domes
97	9	2.1	168	4	Q16256	Q16256 homo sapien	170	7	1.6	136	8	Q9GDH0	Q9gdh0 malus flore
98	8	1.9	59	16	Q932C2	Q932c2 staphylococ	171	7	1.6	136	8	Q9GDG2	Q9gdg2 malus orien
99	8	1.9	116	6	Q9XSQ3	Q9xsq3 lama guanac	172	7	1.6	136	8	Q9GBG1	Q9gdg1 malus pratt
100	8	1.9	124	6	Q95NE6	Q95ne6 bubalus bub	173	7	1.6	136	8	Q9GDF1	Q9gdg1 malus coron
101	8	1.9	142	6	Q9BEC3	Q9bec3 tragulus ja	174	7	1.6	136	8	Q9GDE9	Q9gde9 pyrus salic
102	8	1.9	152	11	Q8VD89	Q8vdb9 rattus norv	175	7	1.6	140	17	Q9VAC0	Q9vac0 aeropyrum p
103	8	1.9	158	10	Q9SYL3	Q9sy13 arabidopsis	176	7	1.6	141	8	Q9GDF4	Q9gdg4 malus trilo
104	8	1.9	226	11	Q9QZD2	Q9qzd2 rattus norv	177	7	1.6	141	8	Q9GDF0	Q9gdg0 pyrus elaea
105	8	1.9	238	17	Q973R9	Q973r9 sulfobus	178	7	1.6	141	8	Q9GDE8	Q9gde8 docynia del
106	8	1.9	278	16	P74248	P74248 synecocyst	179	7	1.6	141	8	Q9GDE7	Q9gde7 cydonia obl
107	8	1.9	288	16	Q92LM2	Q92lm2 rhizobium m	180	7	1.6	151	8	Q9GDE6	Q9gde6 malus hupeh
108	8	1.9	313	12	Q9XVU1	Q9xvu1 melanolpius	181	7	1.6	155	8	Q9GDH3	Q9gdh3 malus domes
109	8	1.9	362	10	Q9ATS8	Q9ats8 cymbopogon	182	7	1.6	155	8	Q9GDH1	Q9gdh1 malus bacca
110	8	1.9	365	16	Q9JQM5	Q9jqm5 neisseria m	183	7	1.6	155	8	Q9GDC8	Q9gdc8 malus fusca
111	8	1.9	366	10	Q9ATS4	Q9ats4 capillipedi	184	7	1.6	155	8	Q9GDF3	Q9gdg3 malus tscho
112	8	1.9	366	10	Q9ATS0	Q9ats0 bothriochlo	185	7	1.6	158	2	Q9E2I3	Q9ezi3 borrelia bu
113	8	1.9	379	16	Q9CFB0	Q9cfb0 lactococcus	186	7	1.6	158	5	Q94577	Q94577 helicoidari
114	8	1.9	416	5	Q96144	Q96144 plasmodium	187	7	1.6	162	17	Q59581	Q59581 pyrococcus
115	8	1.9	431	5	Q9VFP2	Q9vfp2 drosophila	188	7	1.6	168	17	Q9YBR2	Q9ybr2 aeropyrum p
116	8	1.9	503	8	Q8WJQ5	Q8wjq5 gillenia st	189	7	1.6	173	16	Q9PR48	Q9pr48 ureaplasma
117	8	1.9	516	13	Q42186	Q42186 brachydanio	190	7	1.6	176	2	Q9F933	Q9f933 borrelia bu
118	8	1.9	517	13	Q90VL6	Q90vl6 brachydanio	191	7	1.6	176	5	P90669	P90669 aplysia cal
119	8	1.9	560	10	Q94DW0	Q94dw0 oryza sativ	192	7	1.6	178	10	Q9LTP3	Q9ltp3 arabidopsis
120	8	1.9	632	9	Q9FZT9	Q9fzt9 bacterioph	193	7	1.6	185	5	Q00837	Q00837 leishmania
121	8	1.9	636	9	Q8SDT4	Q8sdt4 bacterioph	194	7	1.6	190	17	Q9YAR2	Q9yar2 aeropyrum p
122	8	1.9	636	16	Q931W8	Q931w8 staphylococ	195	7	1.6	194	15	Q89968	Q89968 chimpanzee
123	8	1.9	678	5	Q94736	Q94736 stomoxys ca	196	7	1.6	194	15	Q9YVC7	Q9yvc7 chimpanzee
124	8	1.9	815	10	Q9FZ65	Q9fz65 arabidopsis	197	7	1.6	194	15	Q89973	Q89973 chimpanzee
125	8	1.9	923	13	Q90ZT7	Q90zt7 xenopus lae	198	7	1.6	194	15	Q9YVC6	Q9yvc6 chimpanzee
126	8	1.9	1281	13	Q918L4	Q918l4 brachydanio	199	7	1.6	194	15	Q9YVB8	Q9yvb8 chimpanzee
127	8	1.9	3309	5	Q9GUP2	Q9gup2 caenorhabdi	200	7	1.6	195	16	Q97KX0	Q97kx0 clostridium
128	7	1.6	44	12	Q65825	Q65825 unidentified	201	7	1.6	203	5	Q17346	Q17346 caenorhabdi
129	7	1.6	83	5	Q20689	Q20689 caenorhabdi	202	7	1.6	207	10	Q04307	Q04307 arabidopsis
130	7	1.6	88	16	P96637	P96637 bacillus su	203	7	1.6	209	16	Q45593	Q45593 bacillus su
131	7	1.6	90	17	Q9HJR9	Q9hjr9 thermoplas	204	7	1.6	210	16	Q07036	Q07036 mycobacteri
132	7	1.6	91	9	Q9AYU7	Q9ayu7 bacterioph	205	7	1.6	211	10	Q9MIF8	Q9mif8 arabidopsis
133	7	1.6	91	16	Q9CB09	Q9cb09 lactococcus	206	7	1.6	211	17	Q29130	Q29130 archaeoglob
134	7	1.6	92	9	Q9MCG5	Q9mcg5 bacterioph	207	7	1.6	214	10	Q9SUA6	Q9sua6 arabidopsis
135	7	1.6	92	10	Q64820	Q64820 arabidopsis	208	7	1.6	214	16	Q8VIR3	Q8vir3 mycobacteri
136	7	1.6	99	8	Q9GDF6	Q9gdf6 sorbus torm	209	7	1.6	214	16	Q8RI83	Q8ri83 fusbobacteri
137	7	1.6	101	9	Q92X17	Q92x17 mycobacteri	210	7	1.6	214	16	Q8RF02	Q8rf02 fusbobacteri
138	7	1.6	102	8	Q8SLI7	Q8sl17 physocarpus	211	7	1.6	218	10	Q93WN6	Q93wn6 oryza sativ
139	7	1.6	102	8	Q8SLI6	Q8sl16 physocarpus	212	7	1.6	219	12	Q55579	Q55579 leucania se
140	7	1.6	102	8	Q8SLI5	Q8sl15 neillia aff	213	7	1.6	226	17	Q51760	Q51760 pyrococcus
141	7	1.6	102	8	Q8SLI4	Q8sl14 neillia gra	214	7	1.6	227	3	Q94348	Q94348 schizosacch
142	7	1.6	102	8	Q8SEI2	Q8sei2 neillia thy	215	7	1.6	227	4	Q9NTA4	Q9nta4 homo sapien
143	7	1.6	109	5	Q77411	Q77411 plasmodium	216	7	1.6	227	16	Q8UCQ9	Q8ucq9 agrobacteri
144	7	1.6	109	5	Q9TY36	Q9ty36 plasmodium	217	7	1.6	242	2	Q9L3M4	Q9l3m4 rhizobium l
145	7	1.6	109	10	Q9ZUJ9	Q9zuj9 arabidopsis	218	7	1.6	249	16	Q98Q03	Q98q03 mycoplasma
146	7	1.6	109	11	Q9D6B9	Q9d6b9 mus musculu	219	7	1.6	250	16	Q97DG8	Q97dg8 clostridium
147	7	1.6	112	17	Q58272	Q58272 pyrococcus	220	7	1.6	253	2	Q9EZH3	Q9ezh3 isosphaera
148	7	1.6	120	10	Q9ZUP0	Q9zup0 arabidopsis	221	7	1.6	265	10	Q9LX31	Q9lx31 arabidopsis
149	7	1.6	122	10	Q9C8V8	Q9c8v8 arabidopsis	222	7	1.6	265	10	Q8VZRA	Q8vzra arabidopsis
150	7	1.6	123	8	Q9G0V8	Q9g0v8 malus domes	223	7	1.6	265	10	Q8VZRA	Q8vzra arabidopsis
151	7	1.6	123	8	Q9GDF8	Q9gdf8 malus sieve	224	7	1.6	270	16	Q8RD77	Q8rd77 thermoaer
152	7	1.6	124	8	Q9GDC4	Q9gdc4 malus micro	225	7	1.6	276	10	Q94BC4	Q94bc4 petiveria a
153	7	1.6	125	8	Q9GDF9	Q9gdf9 malus sieve	226	7	1.6	276	10	Q94BL4	Q94bl4 agdestis cl
154	7	1.6	129	8	Q9GDF7	Q9gdf7 malus sieve	227	7	1.6	277	10	Q94BA1	Q94ba1 streptomyc
155	7	1.6	131	8	Q9GPG3	Q9gpg3 malus niedz	228	7	1.6	278	16	Q50494	Q50494 trichodiade
156	7	1.6	133	8	Q9GDF8	Q9gdf8 malus syvie	229	7	1.6	280	10	Q94B86	Q94b86 psilotrichu
157	7	1.6	133	8	Q9GDF7	Q9gdf7 malus sieve	230	7	1.6	280	10	Q94B84	Q94b84 psilotrichu
158	7	1.6	133	8	Q9GDB2	Q9gdb2 malus asiat	231	7	1.6	280	10	Q94B94	Q94b94 suaeda vera
159	7	1.6	133	8	Q9GDC0	Q9gdc0 malus siebo	232	7	1.6	281	10	Q94BF5	Q94bf5 gypsophila
160	7	1.6	133	8	Q9GDF5	Q9gdf5 malus torin	233	7	1.6	281	10	Q94BC8	Q94bc8 oxybaphus n
161	7	1.6	134	6	Q9GM51	Q9gm51 sus scrofa	234	7	1.6	281	10	Q94BC0	Q94bc0 pisonia umb
162	7	1.6	134	8	Q9GDC9	Q9gdc9 malus fusca	235	7	1.6	281	10	Q94BC0	Q94bc0 pisonia umb

236	7	1.6	282	10	Q94BK9	Q94bk9 alternanthe	309	7	1.6	391	5	Q9VKZ1	Q9vkz1 drosophila
237	7	1.6	282	10	Q94BK4	Q94bk4 armeria mar	310	7	1.6	401	2	O54672	O54672 lactococcus
238	7	1.6	283	10	Q94BL1	Q94bl1 alluaudia a	311	7	1.6	401	2	O48702	O48702 lactococcus
239	7	1.6	283	10	Q94BD7	Q94bd7 moehringia	312	7	1.6	401	2	Q93K23	Q93k23 lactococcus
240	7	1.6	284	10	Q94B7	Q94b7 beta tridgyn	313	7	1.6	405	17	Q980X0	Q980y0 sulfolobus
241	7	1.6	284	16	Q97HD7	Q97hd7 clostridium	314	7	1.6	406	16	Q8RA81	Q8ra81 thermoanaer
242	7	1.6	285	10	Q94BE0	Q94be0 silene coro	315	7	1.6	407	5	Q8TI06	Q8ti06 bombyx mori
243	7	1.6	289	4	Q9Y387	Q9y387 homo sapien	316	7	1.6	409	12	Q9QAI3	Q9qai3 murid herpe
244	7	1.6	291	10	Q94BJ0	Q94bj0 calandrinia	317	7	1.6	411	5	Q9SVV1	Q9savl1 drosophila
245	7	1.6	293	10	Q94BL0	Q94bl0 alluaudiops	318	7	1.6	413	5	Q9N8L7	Q9n8l7 trypanosoma
246	7	1.6	294	10	Q94BD4	Q94bd4 nepenthes a	319	7	1.6	418	10	Q9FH56	Q9fh56 arabidopsis
247	7	1.6	294	2	Q93921	Q93921 synecococc	320	7	1.6	418	10	O22995	O22995 arabidopsis
248	7	1.6	297	5	P91314	P91314 caenorhabdi	321	7	1.6	420	3	O42980	O42980 schizosacch
249	7	1.6	302	5	Q9NEG3	Q9neg3 drosophila	322	7	1.6	423	16	Q8RIQ1	Q8riq1 fusobacteri
250	7	1.6	303	2	Q9AH90	Q9ah90 streptococc	323	7	1.6	427	3	P78807	P78807 schizosacch
251	7	1.6	311	5	O76700	O76700 caenorhabdi	324	7	1.6	428	8	Q9SET8	Q9set8 adenia digi
252	7	1.6	313	17	O8TSB9	O8tsb9 methanosarc	325	7	1.6	429	12	P88994	P88994 murid herpe
253	7	1.6	314	10	Q9SPJ7	Q9spj7 gossypium h	326	7	1.6	433	4	Q96CW1	Q96cw1 homo sapien
254	7	1.6	318	3	Q9HEU6	Q9heu6 emericella	327	7	1.6	436	16	Q8X4U5	Q8x4u5 escherichia
255	7	1.6	324	16	Q9X6A9	Q9x6a9 yersinia pe	328	7	1.6	443	13	Q9I863	Q9i863 fugu rubrip
256	7	1.6	327	5	Q9W4V4	Q9w4v4 drosophila	329	7	1.6	447	10	Q9FLD7	Q9fld7 arabidopsis
257	7	1.6	332	2	Q9F7R1	Q9f7r1 uncultured	330	7	1.6	449	5	Q9XZN7	Q9xzn7 asterias ru
258	7	1.6	333	10	Q9S849	Q9s849 arabidopsis	331	7	1.6	459	2	O5I848	O5i848 pseudomonas
259	7	1.6	336	2	Q9EIS0	Q9ey50 lactobacilli	332	7	1.6	459	2	Q9R9S6	Q9r9s6 pseudomonas
260	7	1.6	343	5	O01575	O01575 caenorhabdi	333	7	1.6	459	2	Q9K4S4	Q9k4s4 pseudomonas
261	7	1.6	344	2	P74906	P74906 thermus the	334	7	1.6	459	2	O51743	O51743 pseudomonas
262	7	1.6	348	5	Q9VVM4	Q9vvw4 drosophila	335	7	1.6	459	2	P95566	P95566 pseudomonas
263	7	1.6	348	10	O49457	O49457 arabidopsis	336	7	1.6	459	10	Q9SND7	Q9snd7 arabidopsis
264	7	1.6	349	3	Q9HEB3	Q9heb3 neurospora	337	7	1.6	459	16	Q9R6R3	Q9r6r3 thermoanaer
265	7	1.6	354	5	Q9SSW8	Q9ssw8 dictyosteli	338	7	1.6	465	10	Q9AU43	Q9au45 heuchera ca
266	7	1.6	356	10	Q9ATR4	Q9atr4 oryza sativ	339	7	1.6	466	17	Q9V2D8	Q9v2d8 pyrococcus
267	7	1.6	356	13	Q98UK5	Q98uk5 brachydanio	340	7	1.6	467	13	Q9YGC6	Q9ygc6 xenopus lae
268	7	1.6	356	13	O73679	O73679 brachydanio	341	7	1.6	471	11	Q9Z0M5	Q9z0m5 mus musculu
269	7	1.6	356	16	Q9XEI9	Q9xel9 escherichia	342	7	1.6	473	10	Q9LX54	Q9lx54 arabidopsis
270	7	1.6	357	5	O16144	O16144 ascaris suu	343	7	1.6	476	2	Q9LCK4	Q9lck4 rikkenella m
271	7	1.6	357	16	Q9R5M5	Q9r5m5 fusobacteri	344	7	1.6	479	11	Q9IWP5	Q9iwp5 mus musculu
272	7	1.6	358	10	Q9ATR8	Q9att8 andropogon	345	7	1.6	480	11	Q9IXE1	Q9ixel mus musculu
273	7	1.6	368	16	Q9PHL2	Q9phl2 campylobact	346	7	1.6	481	11	Q9IWT3	Q9iwt3 mus musculu
274	7	1.6	373	16	O34625	O34625 bacillus su	347	7	1.6	481	11	Q9IWT1	Q9iwt1 mus musculu
275	7	1.6	374	5	Q9VNH5	Q9vnh5 drosophila	348	7	1.6	482	11	Q9IX92	Q9ix92 mus musculu
276	7	1.6	375	8	Q9M5C4	Q9m5c4 amaranthus	349	7	1.6	484	5	Q9N487	Q9n487 caenorhabdi
277	7	1.6	376	8	Q9MSG2	Q9msg2 nepenthes v	350	7	1.6	484	11	Q9VEA0	Q9vea0 mus musculu
278	7	1.6	376	8	Q9MSG1	Q9msg1 nepenthes t	351	7	1.6	486	6	Q95L68	Q95l68 macaca fasc
279	7	1.6	376	8	Q9MSG0	Q9msg0 nepenthes t	352	7	1.6	488	11	Q9IWR1	Q9iwr1 mus musculu
280	7	1.6	376	8	Q9MSF9	Q9msf9 nepenthes t	353	7	1.6	489	11	Q9VCX4	Q9vcx4 mus musculu
281	7	1.6	376	8	Q9MSF8	Q9msf8 nepenthes m	354	7	1.6	499	8	Q8WJP0	Q8wjp0 prunus virg
282	7	1.6	376	8	Q9MSF7	Q9msf7 nepenthes v	355	7	1.6	500	8	Q8WJP2	Q8wjp2 prunus laur
283	7	1.6	376	8	Q9MSF6	Q9msf6 nepenthes a	356	7	1.6	500	16	Q9RT87	Q9rt87 deinococcus
284	7	1.6	376	8	Q9MSF5	Q9msf5 nepenthes m	357	7	1.6	501	8	Q95GP4	Q95gp4 nepenthes l
285	7	1.6	376	8	Q9MSF4	Q9msf4 nepenthes k	358	7	1.6	501	8	Q8WJP9	Q8wjp9 neviusia al
286	7	1.6	376	8	Q9MSF3	Q9msf3 nepenthes p	359	7	1.6	502	8	Q8WJN2	Q8wjn2 spiraea can
287	7	1.6	376	8	Q9MSF2	Q9msf2 nepenthes d	360	7	1.6	503	8	Q95GR3	Q95gr3 nepenthes m
288	7	1.6	376	8	Q9MSF1	Q9msf1 anastrocia	361	7	1.6	503	8	Q95GP8	Q95gp8 nepenthes m
289	7	1.6	376	8	Q9MSF0	Q9msf0 anastrocia	362	7	1.6	503	8	Q95E99	Q95e99 lobivia pen
290	7	1.6	376	8	Q9MSE9	Q9mse9 anastrocia	363	7	1.6	503	8	Q95E98	Q95e98 harrisia po
291	7	1.6	376	8	Q9MSE8	Q9mse8 anastrocia	364	7	1.6	503	16	Q9A4M2	Q9a4m2 caulobacter
292	7	1.6	376	8	Q9MSE7	Q9mse7 triphophyll	365	7	1.6	504	6	Q9BDM1	Q9bdl1 pongo pygma
293	7	1.6	376	8	Q9MSE6	Q9mse6 dioncophyll	366	7	1.6	504	8	Q95GV7	Q95gv7 nepenthes g
294	7	1.6	376	8	Q9MSE5	Q9mse5 habropetalu	367	7	1.6	504	8	Q95GV6	Q95gv6 nepenthes b
295	7	1.6	376	8	Q9MSD4	Q9msd4 armeria gad	368	7	1.6	504	8	Q95GV5	Q95gv5 nepenthes a
296	7	1.6	376	8	Q9MSC7	Q9msc7 simmondsia	369	7	1.6	504	8	Q95GV4	Q95gv4 nepenthes e
297	7	1.6	376	8	Q9MSC6	Q9msc6 chenopodium	370	7	1.6	504	8	Q95GV3	Q95gv3 nepenthes s
298	7	1.6	376	8	Q9MSC5	Q9msc5 bougainvill	371	7	1.6	504	8	Q95GV2	Q95gv2 nepenthes d
299	7	1.6	380	2	Q9EZH4	Q9ezh4 pirellula m	372	7	1.6	504	8	Q95GV1	Q95gv1 nepenthes i
300	7	1.6	382	17	Q97W74	Q97w74 sulfolobus	373	7	1.6	504	8	Q95GV0	Q95gv0 nepenthes l
301	7	1.6	385	2	Q48696	Q48696 lactococcus	374	7	1.6	504	8	Q95GU9	Q95gu9 nepenthes o
302	7	1.6	386	2	Q9AC99	Q9ac99 lactococcus	375	7	1.6	504	8	Q95GU8	Q95gu8 nepenthes l
303	7	1.6	388	2	O54685	O54685 lactococcus	376	7	1.6	504	8	Q95GU6	Q95gu6 nepenthes c
304	7	1.6	388	2	Q48697	Q48697 lactococcus	377	7	1.6	504	8	Q95GU5	Q95gu5 nepenthes c
305	7	1.6	388	5	Q95R32	Q95r32 drosophila	378	7	1.6	504	8	Q95GU4	Q95gu4 nepenthes r
306	7	1.6	388	5	Q9VF24	Q9vf24 drosophila	379	7	1.6	504	8	Q95GU3	Q95gu3 nepenthes r
307	7	1.6	388	10	Q94LM8	Q94lm8 oryza sativ	380	7	1.6	504	8	Q95GU2	Q95gu2 nepenthes i
308	7	1.6	389	10	Q91K76	Q91k76 arabidopsis	381	7	1.6	504	8	Q95GU1	Q95gu1 nepenthes i

382	7	1.6	504	8	095G00	095G00 nepenthes m	455	7	1.6	507	8	095EA4	095ea4 brownlingia
383	7	1.6	504	8	095GT9	095gt9 nepenthes m	456	7	1.6	507	8	095EA2	095ea2 coleocephal
384	7	1.6	504	8	095GT8	095gt8 nepenthes p	457	7	1.6	507	8	095EA1	095ea1 uebelmannia
385	7	1.6	504	8	095GT7	095gt7 nepenthes d	458	7	1.6	507	8	095EA0	095ea0 trichocereu
386	7	1.6	504	8	095GT6	095gt6 nepenthes k	459	7	1.6	507	8	095E97	095e97 rauhocereus
387	7	1.6	504	8	095GT5	095gt5 nepenthes a	460	7	1.6	507	8	095E96	095e96 matucana in
388	7	1.6	504	8	095GT4	095gt4 nepenthes h	461	7	1.6	507	8	095E93	095e93 parodia maa
389	7	1.6	504	8	095GT3	095gt3 nepenthes t	462	7	1.6	507	8	095E90	095e90 eriosyce au
390	7	1.6	504	8	095GT2	095gt2 nepenthes a	463	7	1.6	507	8	095E89	095e89 eriosyce is
391	7	1.6	504	8	095GT1	095gt1 nepenthes v	464	7	1.6	507	8	095E88	095e88 eriosyce su
392	7	1.6	504	8	095GT0	095gt0 nepenthes t	465	7	1.6	507	8	095E87	095e87 neowerderma
393	7	1.6	504	8	095GS9	095gs9 nepenthes m	466	7	1.6	507	8	094PY9	094py9 acanthocaly
394	7	1.6	504	8	095GS8	095gs8 nepenthes v	467	7	1.6	507	8	094P94	094p94 eriosyce na
395	7	1.6	504	8	095GS7	095gs7 nepenthes n	468	7	1.6	507	8	094P31	094p31 haageocereu
396	7	1.6	504	8	095GS6	095gs6 nepenthes v	469	7	1.6	507	8	094NC4	094nc4 stetsonia c
397	7	1.6	504	8	095GS5	095gs5 nepenthes t	470	7	1.6	508	8	095EB2	095eb2 pfeiffera m
398	7	1.6	504	8	095GS4	095gs4 nepenthes t	471	7	1.6	508	8	095EB3	095eb3 pereskia gu
399	7	1.6	504	8	095GR9	095gr9 nepenthes t	472	7	1.6	509	8	095EB2	095eb2 pereskia st
400	7	1.6	504	8	095GR8	095gr8 nepenthes l	473	7	1.6	509	8	095EB1	095eb1 pereskia zi
401	7	1.6	504	8	095GR7	095gr7 nepenthes e	474	7	1.6	509	8	095EE0	095ee0 austrocylin
402	7	1.6	504	8	095GR6	095gr6 nepenthes r	475	7	1.6	509	8	095ED9	095ed9 opuntia qui
403	7	1.6	504	8	095GR5	095gr5 nepenthes p	476	7	1.6	509	8	095ED5	095ed5 blossfeldia
404	7	1.6	504	8	095GR4	095gr4 nepenthes m	477	7	1.6	509	8	095ED4	095ed4 frailea gra
405	7	1.6	504	8	095GR1	095gr1 nepenthes d	478	7	1.6	509	8	095ED3	095ed3 frailea pha
406	7	1.6	504	8	095GR0	095gr0 nepenthes f	479	7	1.6	509	8	095ED2	095ed2 calymmanthi
407	7	1.6	504	8	095GQ9	095gq9 nepenthes p	480	7	1.6	509	8	095EC7	095ec7 acanthocereu
408	7	1.6	504	8	095GQ8	095gq8 nepenthes m	481	7	1.6	509	8	095EC3	095ec3 armatocereu
409	7	1.6	504	8	095GQ7	095gq7 nepenthes b	482	7	1.6	509	8	095EC2	095ec2 leptocereu
410	7	1.6	504	8	095GQ6	095gq6 nepenthes s	483	7	1.6	509	8	095EC1	095ec1 castellanos
411	7	1.6	504	8	095GQ5	095gq5 nepenthes s	484	7	1.6	509	8	095EC0	095ec0 neoraimondi
412	7	1.6	504	8	095GQ4	095gq4 nepenthes t	485	7	1.6	509	8	095EB9	095eb9 austrocactu
413	7	1.6	504	8	095GQ3	095gq3 nepenthes v	486	7	1.6	509	8	095EB8	095eb8 eulychinia l
414	7	1.6	504	8	095GQ2	095gq2 nepenthes d	487	7	1.6	509	8	095EB7	095eb7 cortyocactu
415	7	1.6	504	8	095GQ0	095gq0 nepenthes s	488	7	1.6	509	8	095EB6	095eb6 corryocactu
416	7	1.6	504	8	095GP9	095gp9 nepenthes e	489	7	1.6	509	8	095EB5	095eb5 pfeiffera l
417	7	1.6	504	8	095GP5	095gp5 nepenthes f	490	7	1.6	509	8	095EB4	095eb4 pfeiffera p
418	7	1.6	504	8	095GP3	095gp3 nepenthes s	491	7	1.6	509	8	095EB3	095eb3 eschincocreu
419	7	1.6	504	8	095GP2	095gp2 nepenthes g	492	7	1.6	509	8	095EB1	095eb1 lophocereus
420	7	1.6	504	8	094Q55	094q55 nepenthes r	493	7	1.6	509	8	095EA9	095ea9 hylocereus
421	7	1.6	504	8	094PH2	094ph2 nepenthes r	494	7	1.6	509	8	095EA8	095ea8 selenicereu
422	7	1.6	504	8	094NV9	094nv9 nepenthes b	495	7	1.6	509	8	095EA7	095ea7 discocactus
423	7	1.6	504	8	094NR5	094nr5 nepenthes a	496	7	1.6	509	8	095EA6	095ea6 parodia mic
424	7	1.6	504	8	08WJR5	08wjr5 adenostoma	497	7	1.6	509	8	095E92	095e92 parodia ott
425	7	1.6	504	8	08WJR4	08wjr4 aruncus dio	498	7	1.6	509	8	095E91	095e91 hatoriola sal
426	7	1.6	504	8	08WJR1	08wjr1 chamaebatia	499	7	1.6	509	8	095E86	095e86 rhipsalis f
427	7	1.6	504	8	08WJRO	08wjro cotoneaster	500	7	1.6	509	8	095E85	095e85 schlumberge
428	7	1.6	504	8	08WJQ8	08wjq8 exochorda r	501	7	1.6	509	8	095E84	095e84 leptismium c
429	7	1.6	504	8	08WJQ4	08wjq4 gillenla tr	502	7	1.6	509	8	095E83	095e83 grahamia br
430	7	1.6	504	8	08WJQ2	08wjq2 kageneckia	503	7	1.6	510	8	095EE5	095ee5 grahamia br
431	7	1.6	504	8	08WJQ1	08wjq1 lyonothamn	504	7	1.6	510	8	08WKL8	08wkl8 populum nig
432	7	1.6	504	8	08WJQ0	08wjq0 neillia thy	505	7	1.6	510	8	09CKF9	09ckf9 pasteurellia
433	7	1.6	504	8	08WJP8	08wjp8 oemleria ce	506	7	1.6	511	5	08WQO4	08wqo4 calliphora
434	7	1.6	504	8	08WJP7	08wjp7 photinia se	507	7	1.6	511	8	095ED7	095ed7 maihuenia p
435	7	1.6	504	8	08WJP6	08wjp6 physocarpus	508	7	1.6	511	8	095ED6	095ed6 maihuenia p
436	7	1.6	504	8	08WJP4	08wjp4 prinsepia s	509	7	1.6	512	8	095G07	095g07 nepenthes l
437	7	1.6	504	8	08WJN8	08wjn8 pyrus cauca	510	7	1.6	514	8	095GP1	095gp1 ancistrocla
438	7	1.6	504	8	08WJN6	08wjn6 rhodotypos	511	7	1.6	517	5	09VX85	09vx85 drosophila
439	7	1.6	504	8	08WJN4	08wjn4 sorbaria so	512	7	1.6	520	16	08XKT0	08xkt0 clostridium
440	7	1.6	504	8	08WJN3	08wjn3 sorbus cali	513	7	1.6	520	11	09QXU8	09qxu8 rattus norv
441	7	1.6	504	8	08WJN0	08wjn0 vaquelinia	514	7	1.6	524	4	09H188	09h188 homo sapien
442	7	1.6	504	16	092LK8	092lk8 rhizobium m	515	7	1.6	525	4	09H187	09h187 homo sapien
443	7	1.6	505	8	095EA3	095ea3 gymnocalyci	516	7	1.6	525	4	09H186	09h186 homo sapien
444	7	1.6	505	8	095E95	095e95 parodia has	517	7	1.6	525	4	09H185	09h185 homo sapien
445	7	1.6	505	8	095E94	095e94 parodia bue	518	7	1.6	525	4	09H184	09h184 homo sapien
446	7	1.6	506	8	08WJQ9	08wjq9 crataegus m	519	7	1.6	525	4	09H183	09h183 homo sapien
447	7	1.6	506	8	08WJQ3	08wjq3 holodiscus	520	7	1.6	525	4	09H182	09h182 homo sapien
448	7	1.6	506	8	08WJP3	08wjp3 prunus dulc	521	7	1.6	525	4	09H181	09h181 homo sapien
449	7	1.6	506	8	08WJP1	08wjp1 prunus pers	522	7	1.6	525	6	09BG53	09bg53 pan troglod
450	7	1.6	507	8	095ED8	095ed8 peresklopsi	523	7	1.6	528	16	09RTP5	09rtp5 deinococcus
451	7	1.6	507	8	095EC6	095ec6 coplapoa br	524	7	1.6	529	5	095Z26	095z26 caenorhabdi
452	7	1.6	507	8	095EC5	095ec5 coplapoa la	525	7	1.6	531	3	04Z797	04z797 podospora a
453	7	1.6	507	8	095EC4	095ec4 coplapoa la	526	7	1.6	532	5	017500	017500 caenorhabdi
454	7	1.6	507	8	095EA5	095ea5 brownlingia	527	7	1.6	532	5		

528	7	1.6	538	10	Q9XFM4	Q9xfm4 fagopyrum e	601	7	1.6	905	13	Q90505	Q90505 fundulus he
529	7	1.6	539	10	Q94LS6	Q94ls6 oryza sativ	602	7	1.6	907	16	Q8y0t5	Q8y0t5 ralstonia s
530	7	1.6	540	10	Q08737	Q08737 triticum ae	603	7	1.6	913	16	Q9PQ03	Q9pq03 ureaplasma
531	7	1.6	550	3	Q9P5P9	Q9p5p9 neurospora	604	7	1.6	929	5	Q9NGW5	Q9ngw5 drosophila
532	7	1.6	553	3	Q9WQp6	Q9wqp6 drosophila	605	7	1.6	929	5	Q9NBW0	Q9nbw0 drosophila
533	7	1.6	554	2	Q9ZFF1	Q9zff1 klebsiella	606	7	1.6	929	5	Q9NBL3	Q9nbl3 drosophila
534	7	1.6	554	2	Q59470	Q59470 klebsiella	607	7	1.6	938	10	Q9LDD7	Q9ldd7 lycopersico
535	7	1.6	554	16	Q825M2	Q825m2 salmonella	608	7	1.6	939	5	Q9NHQ0	Q9nhq0 drosophila
536	7	1.6	555	5	Q16745	Q16745 caenorhabdi	609	7	1.6	950	5	Q95TRA	Q95tr4 drosophila
537	7	1.6	573	16	Q926K8	Q926k8 listeria in	610	7	1.6	951	10	Q9AR52	Q9ar52 vicia faba
538	7	1.6	579	2	Q93LM7	Q93lm7 pseudomonas	611	7	1.6	951	10	Q43106	Q43106 phaseolus v
539	7	1.6	587	10	Q9ZWF9	Q9zwf9 cucumis mel	612	7	1.6	956	10	Q43131	Q43131 vicia faba
540	7	1.6	588	5	Q9NCM4	Q9ncm4 plasmodium	613	7	1.6	956	10	Q42932	Q42932 nicotiana p
541	7	1.6	588	5	Q9NCM2	Q9ncm2 plasmodium	614	7	1.6	956	10	Q43182	Q43182 solanum tub
542	7	1.6	595	5	Q9NCM1	Q9ncm1 plasmodium	615	7	1.6	956	10	Q9SPD5	Q9spd5 lycopersico
543	7	1.6	595	5	Q9NCM0	Q9ncm0 plasmodium	616	7	1.6	957	10	Q43002	Q43002 oryza sativ
544	7	1.6	596	5	Q19953	Q19953 caenorhabdi	617	7	1.6	958	4	Q96A52	Q96a52 homo sapien
545	7	1.6	597	5	Q9N657	Q9n657 plasmodium	618	7	1.6	974	5	Q20143	Q20143 caenorhabdi
546	7	1.6	597	5	Q9NCN3	Q9ncn3 plasmodium	619	7	1.6	989	10	Q9SCM5	Q9scm5 arabidopsis
547	7	1.6	597	5	Q9NCN1	Q9ncn1 plasmodium	620	7	1.6	1014	12	Q8QS56	Q8qs56 chimpanzee
548	7	1.6	597	5	Q9NCM6	Q9ncm6 plasmodium	621	7	1.6	1019	5	Q77143	Q77143 drosophila
549	7	1.6	597	5	Q9NCM3	Q9ncm3 plasmodium	622	7	1.6	1021	5	Q77289	Q77289 drosophila
550	7	1.6	599	5	Q9NCN2	Q9ncn2 plasmodium	623	7	1.6	1024	5	Q9W268	Q9w268 drosophila
551	7	1.6	599	5	Q9NCM7	Q9ncm7 plasmodium	624	7	1.6	1054	11	Q9JM99	Q9jlm99 mus musculu
552	7	1.6	599	5	Q9NCM5	Q9ncm5 plasmodium	625	7	1.6	1063	3	Q96WN2	Q96wn2 paracoccidi
553	7	1.6	609	13	Q9W6M5	Q9w6m5 agkistrodon	626	7	1.6	1083	5	Q8T3K8	Q8t3k8 drosophila
554	7	1.6	610	12	Q71123	Q71123 rhesus cyto	627	7	1.6	1104	16	Q51934	Q51934 thermotoga
555	7	1.6	613	5	Q9NCN0	Q9ncn0 plasmodium	628	7	1.6	1122	5	Q23682	Q23682 caenorhabdi
556	7	1.6	614	5	Q9NCM9	Q9ncm9 plasmodium	629	7	1.6	1161	5	Q9W2X8	Q9w2x8 drosophila
557	7	1.6	616	5	Q9NCM8	Q9ncm8 plasmodium	630	7	1.6	1173	12	Q98177	Q98177 molluscum c
558	7	1.6	624	16	Q99Q28	Q99q28 staphylococ	631	7	1.6	1175	12	Q85285	Q85285 molluscum c
559	7	1.6	630	2	Q9EVD6	Q9evd6 actinomyces	632	7	1.6	1180	5	Q24163	Q24163 drosophila
560	7	1.6	638	5	Q9VYX9	Q9vyx9 drosophila	633	7	1.6	1206	5	Q9VED3	Q9ved3 drosophila
561	7	1.6	642	16	Q9JYK7	Q9jyk7 neisseria m	634	7	1.6	1286	5	Q9U4W1	Q9u4w1 aedes aegypt
562	7	1.6	642	16	Q9JTK6	Q9jtk6 neisseria m	635	7	1.6	1214	10	Q9LD60	Q9ld60 arabidopsis
563	7	1.6	659	5	Q17792	Q17792 caenorhabdi	636	7	1.6	1263	4	Q15044	Q15044 homo sapien
564	7	1.6	666	2	Q32495	Q32495 bacteroides	637	7	1.6	1283	5	Q95RH4	Q95rh4 drosophila
565	7	1.6	666	5	Q9VP48	Q9vp48 drosophila	638	7	1.6	1284	5	Q9VJ79	Q9vj79 drosophila
566	7	1.6	670	5	Q9N554	Q9n554 caenorhabdi	639	7	1.6	1308	5	Q01924	Q01924 caenorhabdi
567	7	1.6	684	2	Q65990	Q65990 clostridium	640	7	1.6	1339	5	Q26048	Q26048 pacifastacu
568	7	1.6	684	16	Q97MN8	Q97mn8 clostridium	641	7	1.6	1441	5	Q96957	Q96957 drosophila
569	7	1.6	685	4	Q96CV7	Q96cv7 homo sapien	642	7	1.6	1446	5	Q77063	Q77063 aplysia cal
570	7	1.6	688	10	Q9S8Y0	Q9s8y0 atropa bell	643	7	1.6	1461	5	Q8SQ03	Q8sq03 encephalito
571	7	1.6	700	2	Q34003	Q34003 rhodobacter	644	7	1.6	1465	5	Q9VMD0	Q9vmd0 drosophila
572	7	1.6	706	5	Q9V417	Q9v417 drosophila	645	7	1.6	1521	5	Q8SX83	Q8sx83 drosophila
573	7	1.6	706	5	Q8SXG1	Q8sxg1 drosophila	646	7	1.6	1534	5	Q9W3D3	Q9w3d3 drosophila
574	7	1.6	718	5	Q9VRH4	Q9vrh4 drosophila	647	7	1.6	1561	5	Q9W3D2	Q9w3d2 drosophila
575	7	1.6	737	16	Q985Z1	Q985z1 fusobacteri	648	7	1.6	1591	3	Q9HFW1	Q9hfw1 ashbya goss
576	7	1.6	758	5	Q96037	Q96037 clona savig	649	7	1.6	1597	5	Q61346	Q61346 drosophila
577	7	1.6	768	16	Q9YANO	Q9yan0 listeria mo	650	7	1.6	1625	10	Q08367	Q08367 zea mays (m
578	7	1.6	784	6	Q9TUN5	Q9tun5 sus scrofa	651	7	1.6	1645	5	Q8WSS9	Q8wss9 drosophila
579	7	1.6	784	6	Q95JH1	Q95jh1 sus scrofa	652	7	1.6	1677	5	Q9BKV5	Q9bkv5 leishmania
580	7	1.6	792	16	Q9PE24	Q9pez4 xylella fas	653	7	1.6	1685	10	Q43248	Q43248 zea mays (m
581	7	1.6	799	17	Q8TH11	Q8th11 pyrococcus	654	7	1.6	1726	5	Q02569	Q02569 plasmodium
582	7	1.6	816	5	Q26026	Q26026 plasmodium	655	7	1.6	1751	5	Q26194	Q26194 plasmodium
583	7	1.6	818	13	Q9PSV9	Q9psv9 xenopus lae	656	7	1.6	1895	5	Q9VWS5	Q9vws5 drosophila
584	7	1.6	818	13	Q91742	Q91742 xenopus lae	657	7	1.6	1911	5	Q9W4M7	Q9w4m7 drosophila
585	7	1.6	828	13	Q9DQK3	Q9dqk3 xenopus lae	658	7	1.6	2038	5	Q967Y0	Q967y0 dictyosteli
586	7	1.6	828	13	Q91743	Q91743 xenopus lae	659	7	1.6	2091	5	Q9VJ69	Q9vj69 drosophila
587	7	1.6	829	13	Q9PSV8	Q9psv8 xenopus lae	660	7	1.6	2280	5	Q9V8E6	Q9v8e6 drosophila
588	7	1.6	832	15	Q93258	Q93258 chimpanzee	661	7	1.6	2302	5	Q9N693	Q9n693 drosophila
589	7	1.6	846	2	Q53112	Q53112 mycoplasma	662	7	1.6	2310	5	Q9GRA9	Q9gra9 drosophila
590	7	1.6	846	2	Q92FP7	Q92fp7 vibrio chol	663	7	1.6	2311	10	Q48959	Q48959 trititum ae
591	7	1.6	851	13	Q9KQP6	Q9kqp6 vibrio chol	664	7	1.6	2321	10	Q947M6	Q947m6 setaria ita
592	7	1.6	851	13	Q42507	Q42507 xenopus lae	665	7	1.6	2325	10	Q41743	Q41743 zea mays (m
593	7	1.6	856	5	Q8SSV0	Q8ssv0 dictyosteli	666	7	1.6	3080	5	Q9VRY3	Q9vry3 drosophila
594	7	1.6	861	5	Q9VRT8	Q9vrt8 drosophila	667	7	1.6	5476	5	Q9NJ17	Q9nj17 drosophila
595	7	1.6	865	16	Q8XY19	Q8xy19 ralstonia s	668	7	1.6	5533	5	Q9VPL2	Q9vpl2 drosophila
596	7	1.6	869	4	Q15572	Q15572 homo sapien	669	7	1.6	5533	5	Q9U6C3	Q9u6c3 drosophila
597	7	1.6	876	10	Q9FZB0	Q9fzb0 arabidopsis	670	7	1.6	5554	5	Q9NHN1	Q9nhn1 drosophila
598	7	1.6	885	16	Q92UL6	Q92ul6 rhizobium m	671	7	1.6	5560	5	Q9VPL1	Q9vpl1 drosophila
599	7	1.6	885	16	Q8XTK5	Q8xtk5 ralstonia s	672	7	1.6	5616	2	Q48439	Q48439 klebsiella
600	7	1.6	897	16	Q98Q69	Q98q69 mycoplasma	673	7	1.4	16	16	Q8X8T5	Q8x8t5 escherichia

674	6	1.4	29	4	Q9UM98	Q9um98 homo sapien	747	6	1.4	89	13	093572	Q93572 ambystoma m
675	6	1.4	29	6	Q9TRL6	Q9trl6 bos taurus	748	6	1.4	92	5	Q18410	Q18410 caenorhabdi
676	6	1.4	32	16	Q9KKK1	Q9kkk1 vibrio chol	749	6	1.4	93	10	Q9SL66	Q9sl66 arabidopsis
677	6	1.4	38	4	Q16097	Q16097 homo sapien	750	6	1.4	93	10	Q9L2N9	Q9l2n9 arabidopsis
678	6	1.4	39	4	Q16098	Q16098 homo sapien	751	6	1.4	93	16	Q926J5	Q926j5 rickettsia
679	6	1.4	46	17	Q8U113	Q8u113 pyrococcus	752	6	1.4	94	2	Q9S139	Q9s139 escherichia
680	6	1.4	48	10	Q8SAS1	Q8sasi pinus sylve	753	6	1.4	94	4	Q14814	Q14814 homo sapien
681	6	1.4	48	10	Q8SAS1	Q8sas1 pinus sylve	754	6	1.4	94	5	Q9U709	Q9u709 plasmodium
682	6	1.4	50	12	Q996F7	Q996f7 human echov	755	6	1.4	94	5	Q9U708	Q9u708 plasmodium
683	6	1.4	51	4	Q96BA5	Q96ba5 homo sapien	756	6	1.4	94	5	Q9U706	Q9u706 plasmodium
684	6	1.4	51	5	Q26845	Q26845 trypanosoma	757	6	1.4	94	5	Q9U705	Q9u705 plasmodium
685	6	1.4	51	5	Q9W582	Q9w5b2 drosophila	758	6	1.4	95	5	Q9U712	Q9u712 plasmodium
686	6	1.4	52	17	Q9P2P0	Q9p2p0 sulfolobus	759	6	1.4	95	5	Q9U711	Q9u711 plasmodium
687	6	1.4	54	4	P78495	P78495 homo sapien	760	6	1.4	95	5	Q9U710	Q9u710 plasmodium
688	6	1.4	54	16	Q8X3B5	Q8x3b5 escherichia	761	6	1.4	95	5	Q9U707	Q9u707 plasmodium
689	6	1.4	55	17	Q8ZWU5	Q8zwu5 pyrobaculum	762	6	1.4	95	5	Q9U704	Q9u704 plasmodium
690	6	1.4	56	5	Q8T2R1	Q8t2r1 dictyosteli	763	6	1.4	95	15	Q88340	Q88340 simian t-ly
691	6	1.4	56	16	Q8YVT1	Q8ytt1 anabaena sp	764	6	1.4	95	15	Q88342	Q88342 simian t-ly
692	6	1.4	57	16	Q9PNN4	Q9pnn4 campylobact	765	6	1.4	96	2	Q33423	Q33423 pseudomonas
693	6	1.4	59	2	Q9RQF7	Q9rqf7 vibrio chol	766	6	1.4	96	5	O61170	O61170 plasmodium
694	6	1.4	59	9	Q9MBU8	Q9mbu8 vibrio phag	767	6	1.4	97	13	Q9W730	Q9w730 acipenser r
695	6	1.4	60	5	Q9VSB4	Q9vsb4 drosophila	768	6	1.4	97	17	Q976D2	Q976d2 sulfolobus
696	6	1.4	60	16	Q8Z523	Q8zs23 anabaena sp	769	6	1.4	99	2	Q97070	Q97070 burkholderi
697	6	1.4	60	16	Q8VK73	Q8vk73 mycobacteri	770	6	1.4	99	2	Q8RP91	Q8rp91 cylindrospe
698	6	1.4	60	17	Q9HMB5	Q9hmb5 halobacteri	771	6	1.4	99	17	Q97239	Q97239 sulfolobus
699	6	1.4	61	12	Q8VAM4	Q8vaw4 white spot	772	6	1.4	100	10	Q8RUF0	Q8ruf0 oryza sativ
700	6	1.4	63	5	Q9VDX0	Q9vdx0 drosophila	773	6	1.4	101	16	Q9FBZ6	Q9fbz6 streptomyce
701	6	1.4	63	16	Q8R7R7	Q8r7r7 thermomanaer	774	6	1.4	102	5	Q94189	Q94189 caenorhabdi
702	6	1.4	64	10	Q40049	Q40049 hordeum vul	775	6	1.4	102	5	Q9VUE1	Q9vuel drosophila
703	6	1.4	64	13	Q91220	Q91220 oncorhynch	776	6	1.4	102	5	Q9VED9	Q9ved9 drosophila
704	6	1.4	67	13	Q92165	Q92165 oncorhynch	777	6	1.4	102	8	Q8SD22	Q8sd22 neillia thi
705	6	1.4	67	13	Q91235	Q91235 oncorhynch	778	6	1.4	102	8	Q8SD21	Q8sd21 neillia thi
706	6	1.4	67	13	Q91172	Q91172 oncorhynch	779	6	1.4	102	8	Q8SD20	Q8sd20 neillia uek
707	6	1.4	67	13	Q91158	Q91158 oncorhynch	780	6	1.4	102	8	Q8SDY9	Q8sdy9 stephanandr
708	6	1.4	67	13	Q92166	Q92166 oncorhynch	781	6	1.4	102	8	Q8SDY8	Q8sdy8 stephanandr
709	6	1.4	69	10	Q8S2R7	Q8s2r7 thellungiel	782	6	1.4	102	8	Q8SDY7	Q8sdy7 stephanandr
710	6	1.4	69	13	Q9P765	Q9pt65 xenopus lae	783	6	1.4	102	10	Q9LVC7	Q9lvc7 arabidopsis
711	6	1.4	69	13	Q9T137	Q9t137 scyllorhinu	784	6	1.4	102	11	P97294	P97294 mus musculus
712	6	1.4	70	5	Q9V374	Q9vv37 drosophila	785	6	1.4	102	16	Q92J02	Q92j02 rickettsia
713	6	1.4	70	16	Q97ME7	Q97me7 clostridium	786	6	1.4	103	15	Q91084	Q91084 human immun
714	6	1.4	71	5	Q9W006	Q9w0q6 drosophila	787	6	1.4	103	17	Q970X8	Q970x8 sulfolobus
715	6	1.4	71	17	Q8TKS8	Q8tk88 methanosarc	788	6	1.4	106	2	Q9KIT8	Q9kit8 bacillus me
716	6	1.4	72	5	Q8T1A1	Q8t1a1 dictyosteli	789	6	1.4	106	2	Q9KIT4	Q9kit4 clostridium
717	6	1.4	72	16	Q25973	Q25973 helicobacte	790	6	1.4	106	10	Q9LS87	Q9ls87 arabidopsis
718	6	1.4	73	5	Q8T844	Q8t844 dictyosteli	791	6	1.4	107	5	Q9VQW5	Q9vqw5 drosophila
719	6	1.4	73	12	Q9WHW7	Q9whw7 tt virus. h	792	6	1.4	108	2	Q82983	Q82983 bacillus sp
720	6	1.4	75	5	Q8T3D9	Q8t3d9 caenorhabdi	793	6	1.4	108	2	Q9XD75	Q9xd75 nitrobacter
721	6	1.4	75	13	Q91456	Q91456 salvelinus	794	6	1.4	108	2	O8VSW4	O8vsw4 staphylococ
722	6	1.4	76	5	Q9VLH4	Q9vlh4 drosophila	795	6	1.4	108	16	Q8XXT6	Q8xxt6 ralstonia s
723	6	1.4	76	16	Q92FX4	Q92fx4 rickettsia	796	6	1.4	109	5	Q9W346	Q9w346 drosophila
724	6	1.4	76	17	Q8U1J9	Q8u1j9 pyrococcus	797	6	1.4	110	12	Q84468	Q84468 paramecium
725	6	1.4	77	5	Q20690	Q20690 caenorhabdi	798	6	1.4	110	17	O58026	O58026 pyrococcus
726	6	1.4	77	16	Q97HT2	Q97ht2 clostridium	799	6	1.4	111	5	O21111	O21111 caenorhabdi
727	6	1.4	77	16	Q9ZJ18	Q9zj18 helicobacte	800	6	1.4	113	16	Q9JX91	Q9jx91 neisseria m
728	6	1.4	78	5	P90569	P90569 plasmodium	801	6	1.4	113	16	Q9JX91	Q9jx91 neisseria m
729	6	1.4	78	16	Q92DV6	Q92dv6 listeria in	802	6	1.4	114	5	Q18724	Q18724 caenorhabdi
730	6	1.4	78	16	Q9Y943	Q9y943 listeria mo	803	6	1.4	115	11	O08932	O08932 mus musculus
731	6	1.4	78	17	Q8TH48	Q8th48 methanopyru	804	6	1.4	115	17	Q9YAJ2	Q9yaj2 aeropyrum p
732	6	1.4	79	4	Q96P11	Q96p11 homo sapien	805	6	1.4	116	5	Q95242	Q95242 plasmodium
733	6	1.4	79	15	Q80853	Q80853 human t-lym	806	6	1.4	116	6	Q97933	Q97933 phocenoide
734	6	1.4	79	15	Q80854	Q80854 human t-lym	807	6	1.4	116	6	Q97934	Q97934 pseudorca c
735	6	1.4	79	15	Q80855	Q80855 human t-lym	808	6	1.4	116	11	Q8VD99	Q8vd99 rattus ratt
736	6	1.4	80	17	Q26182	Q26182 methanobact	809	6	1.4	116	16	Q9K8E3	Q9k8e3 bacillus ha
737	6	1.4	82	10	O80460	O80460 arabidopsis	810	6	1.4	117	11	Q9D1B5	Q9d1b5 mus musculus
738	6	1.4	83	12	Q89074	Q89074 variola vir	811	6	1.4	118	2	O87097	O87097 bacillus su
739	6	1.4	84	5	Q917M5	Q917m5 drosophila	812	6	1.4	118	5	Q9U4E9	Q9u4e9 aedes aegyp
740	6	1.4	84	10	Q9XFE2	Q9xfu2 chlamydomon	813	6	1.4	119	5	Q9NU09	Q9nu09 aedes aegyp
741	6	1.4	84	16	Q97G10	Q97g10 clostridium	814	6	1.4	119	8	Q9ME92	Q9me92 oenothera h
742	6	1.4	86	5	Q9XW64	Q9xw64 caenorhabdi	815	6	1.4	119	16	Q92CB5	Q92cb5 listeria in
743	6	1.4	87	2	Q9APV1	Q9apv1 pseudomonas	816	6	1.4	119	16	O92621	O92621 chlamydia p
744	6	1.4	87	5	Q9I810	Q9i810 drosophila	817	6	1.4	121	9	O8W656	O8w656 bacterioph
745	6	1.4	87	16	Q9I4Y3	Q9i4y3 pseudomonas	818	6	1.4	121	17	Q8ZV63	Q8zv63 pyrobaculum
746	6	1.4	88	5	Q9V5H9	Q9v5h9 drosophila	819	6	1.4	122	5	Q25669	Q25669 plasmodium

820	6	1.4	122	5	Q35670	Q25670 plasmodium	893	6	1.4	143	3	Q94656	O94656 schizosacch
821	6	1.4	122	5	Q25671	Q25671 plasmodium	894	6	1.4	143	5	Q9GQM4	Q9gqm4 trichoplusi
822	6	1.4	122	5	Q9VP82	Q9vp82 drosophila	895	6	1.4	143	9	O64272	O64272 lactococcus
823	6	1.4	122	10	O9FJC5	O9fjc5 arabidopsis	896	6	1.4	143	9	Q38111	Q38111 bacterioph
824	6	1.4	122	10	O23089	O23089 arabidopsis	897	6	1.4	143	10	Q41658	Q41658 vicia faba
825	6	1.4	122	16	O9HXX1	Q9hxx1 pseudomonas	898	6	1.4	143	10	Q9ZWA3	Q9zwa3 arabidopsis
826	6	1.4	122	17	O96Z09	Q96z09 sulfolobus	899	6	1.4	143	11	O60433	O60433 cricetidae
827	6	1.4	123	4	Q96M43	Q96m43 homo sapien	900	6	1.4	143	17	O97XZ5	O97xz5 sulfolobus
828	6	1.4	123	5	Q9VE70	Q9ve70 drosophila	901	6	1.4	143	17	O8TY77	O8ty77 methanopyru
829	6	1.4	124	5	Q9VG13	Q9vg13 drosophila	902	6	1.4	144	8	Q9APQ6	Q9apq6 thrips taba
830	6	1.4	124	6	O9XS40	Q9xs40 camelus bac	903	6	1.4	144	17	Q9HS08	Q9hsd8 halobacteri
831	6	1.4	124	16	Q9R989	Q9r989 thermocanaer	904	6	1.4	145	13	Q9OZ07	Q9ozq7 oryzias lat
832	6	1.4	125	10	Q945S8	Q945s8 arabidopsis	905	6	1.4	146	10	O23043	O23043 arabidopsis
833	6	1.4	126	2	Q9RZN1	Q9rzn1 pseudomonas	906	6	1.4	146	13	Q9I891	Q9i891 gallus gall
834	6	1.4	126	12	Q9EMW9	Q9emw9 amsacta moo	907	6	1.4	146	17	P81332	P81332 methanococc
835	6	1.4	126	16	Q97MR9	Q97mr9 clostridium	908	6	1.4	147	5	Q9U6N0	Q9u6n0 helicoverpa
836	6	1.4	126	16	O9BIL1	Q9bil1 giardia lam	909	6	1.4	148	5	Q9Y038	Q9y038 helicoverpa
837	6	1.4	127	5	Q33968	Q23968 drosophila	910	6	1.4	149	5	Q90019	Q90019 plasmodium
838	6	1.4	128	5	Q9VME4	Q9vme4 drosophila	911	6	1.4	149	5	Q9VA16	Q9vai6 drosophila
839	6	1.4	128	10	Q9MSM4	Q9msm4 euphorbia e	912	6	1.4	149	8	Q8SJP0	Q8sjp0 cynolebias
840	6	1.4	128	16	O9K1A7	Q9k1a7 neisseria m	913	6	1.4	149	16	O8RH25	O8rh25 fusobacteri
841	6	1.4	128	16	O9ZGY3	Q9zgy3 versinia pe	914	6	1.4	149	17	O8TW19	O8tw19 methanopyru
842	6	1.4	128	16	O8RHE6	O8rhe6 fusobacteri	915	6	1.4	150	10	O9FJF8	O9fje8 arabidopsis
843	6	1.4	129	5	Q9VQP9	Q9vqp9 drosophila	916	6	1.4	150	15	O8Q541	O8q541 chimpanzee
844	6	1.4	130	5	Q94208	Q94208 caenorhabdi	917	6	1.4	150	15	O8Q539	O8q539 chimpanzee
845	6	1.4	131	5	Q9NDH9	Q9ndh9 plasmodium	918	6	1.4	150	15	O8Q510	O8q510 chimpanzee
846	6	1.4	131	16	O9Z268	Q9z268 rhizobium m	919	6	1.4	150	15	O8Q4N2	O8q4n2 chimpanzee
847	6	1.4	131	16	O9ZHP2	Q9zhp2 rickettsia	920	6	1.4	150	16	P72786	P72786 synechocyst
848	6	1.4	132	5	O25507	Q25507 manduca sex	921	6	1.4	150	16	O9HV99	O9hvv9 pseudomonas
849	6	1.4	132	13	O9XY97	O9xy97 lampetra fl	922	6	1.4	150	16	O8XK18	O8xk18 clostridium
850	6	1.4	133	2	P95330	P95330 myxococcus	923	6	1.4	151	5	P91600	P91600 chironomus
851	6	1.4	133	16	O8YML1	O8ymll anabaena sp	924	6	1.4	151	5	P91592	P91592 chironomus
852	6	1.4	135	3	O13568	O13568 saccharomyc	925	6	1.4	151	8	Q9G879	Q9g879 malawimonas
853	6	1.4	135	6	O95LJ6	O95lj6 macaca fasc	926	6	1.4	151	10	O9FF22	O9ff22 arabidopsis
854	6	1.4	135	16	O9CC10	O9cc10 mycobacteri	927	6	1.4	151	11	O9D3F1	O9d3f1 mus musculu
855	6	1.4	136	5	Q9GTN6	Q9gtn6 plasmodium	928	6	1.4	151	11	O99P01	O99p01 mus musculu
856	6	1.4	136	5	O93827	Q93827 caenorhabdi	929	6	1.4	152	6	O8SQ10	O8sq10 macaca neme
857	6	1.4	136	11	O08657	O08657 rattus norv	930	6	1.4	152	6	O8SQ09	O8sq09 papio hamad
858	6	1.4	136	17	O9HI89	Q9hi89 thermoplas	931	6	1.4	152	6	O8SPN5	O8spn5 macaca mula
859	6	1.4	137	2	O54029	O54029 propionigen	932	6	1.4	152	6	O8SPF2	O8spf2 cercopithe
860	6	1.4	137	2	P74924	P74924 thermotoga	933	6	1.4	152	10	O9FY58	O9fy58 arabidopsis
861	6	1.4	137	8	O02699	O02699 podospora a	934	6	1.4	152	12	O9SGX9	O9sgx9 helicoverp
862	6	1.4	137	16	O92B88	Q92b88 listeria in	935	6	1.4	152	12	O8V5S7	O8vs57 helicoverpa
863	6	1.4	137	16	O8Z112	Q8z112 versinia pe	936	6	1.4	152	16	Q9X297	Q9x297 thermotoga
864	6	1.4	137	16	O8Y6R3	Q8y6r3 listeria mo	937	6	1.4	153	4	O16869	O16869 homo sapien
865	6	1.4	137	17	O27354	O27354 methanobact	938	6	1.4	153	9	Q8W748	Q8w748 bacterioph
866	6	1.4	137	17	O9Y9B9	Q9y9b9 aeropyrum p	939	6	1.4	153	16	O97FY9	O97fy9 clostridium
867	6	1.4	138	8	O8B3Y2	O8b3y2 terralonus	940	6	1.4	153	16	O8UGY1	O8ugy1 agrobacteri
868	6	1.4	138	8	Q9B3W9	Q9b3w9 paraphidipp	941	6	1.4	154	16	O9KMB9	O9kmb9 vibrio chol
869	6	1.4	138	12	Q8VAS2	Q8vas2 white spot	942	6	1.4	154	16	O98PV6	O98pv6 mycoplasma
870	6	1.4	138	16	O9K098	Q9k098 neisseria m	943	6	1.4	154	17	O97B57	O97b57 thermoplas
871	6	1.4	138	16	O8YXB9	O8yxb9 anabaena sp	944	6	1.4	155	13	O9W732	O9w732 acipenser r
872	6	1.4	138	16	O98R53	O98r53 mycoplasma	945	6	1.4	155	16	O9CJAI	O9cjal lactococcus
873	6	1.4	138	17	O27072	O27072 methanobact	946	6	1.4	155	17	O8U4H5	O8u4h5 pyrococcus
874	6	1.4	139	5	O9Y126	O9y126 hydra magni	947	6	1.4	156	2	O9AKK1	O9akk1 rickettsia
875	6	1.4	139	5	O9UAV7	Q9uav7 caenorhabdi	948	6	1.4	156	2	O9AKK6	O9akk6 rickettsia
876	6	1.4	140	3	P87165	P87165 saccharomyc	949	6	1.4	156	6	O8SQ13	O8sq13 gorilla gor
877	6	1.4	140	5	O26056	O26056 plasmodium	950	6	1.4	156	6	O8SQ12	O8sq12 pongo pygma
878	6	1.4	140	5	O17821	Q17821 caenorhabdi	951	6	1.4	156	6	O8SQ11	O8sq11 hylobates l
879	6	1.4	140	9	O9AZ68	Q9az68 lactococcus	952	6	1.4	156	6	O8SQ08	O8sq08 salmuri sci
880	6	1.4	140	9	O9AZT6	Q9azt6 bacterioph	953	6	1.4	156	6	O8SQ07	O8sq07 saguinus oe
881	6	1.4	140	9	O9AZM9	Q9azm9 bacterioph	954	6	1.4	156	6	O8SQ06	O8sq06 ateles geof
882	6	1.4	140	10	O8RZM6	Q8rz6 oryza sativ	955	6	1.4	156	6	O8SQ05	O8sq05 lagotrix l
883	6	1.4	140	16	O9CIA0	O9cia0 lactococcus	956	6	1.4	156	6	O8SPN4	O8spn4 pygathrix n
884	6	1.4	140	16	O9CFR2	O9cfr2 lactococcus	957	6	1.4	156	6	O8SPN3	O8spn3 pygathrix n
885	6	1.4	141	5	O01816	O01816 caenorhabdi	958	6	1.4	156	8	Q9G4E0	Q9g4e0 thraustochy
886	6	1.4	141	10	Q9LS09	Q9lsq9 arabidopsis	959	6	1.4	156	11	Q9D5F7	Q9d5f7 mus musculu
887	6	1.4	141	16	O9ZF27	Q9zf27 listeria mo	960	6	1.4	156	16	Q92GN3	Q92gn3 rickettsia
888	6	1.4	141	16	O8U6P1	O8u6p1 agrobacteri	961	6	1.4	157	13	Q9PVJ2	Q9pvj2 charina tri
889	6	1.4	142	2	O44391	O44391 agrobacteri	962	6	1.4	157	16	O9KHE7	O9khe7 anabaena sp
890	6	1.4	142	9	Q38384	Q38384 bacterioph	963	6	1.4	158	2	O52404	O52404 edwardsiell
891	6	1.4	142	12	O8U216	Q8uz16 phococoea sp	964	6	1.4	158	17	O8U3W5	O8u3w5 pyrococcus
892	6	1.4	142	13	O13032	O13032 brachydanio	965	6	1.4	159	5	Q9VQR2	Q9vqr2 drosophila

966 1.4 159 10 Q9XG40
 967 6 1.4 159 12 Q9YMT6
 968 6 1.4 160 2 Q9KI53
 969 6 1.4 161 3 Q12155
 970 6 1.4 161 4 Q9NWT0
 971 6 1.4 161 4 Q8WZ09
 972 6 1.4 161 16 Q9BBL0
 973 6 1.4 161 16 Q8R987
 974 6 1.4 162 16 Q9KDR7
 975 6 1.4 162 16 Q97Q11
 976 6 1.4 163 2 Q9ZAR6
 977 6 1.4 163 8 Q9BSC2
 978 6 1.4 163 8 Q9T9N9
 979 6 1.4 163 16 Q9JVA7
 980 6 1.4 163 17 Q58287
 981 6 1.4 165 5 Q02457
 982 6 1.4 165 8 Q03961
 983 6 1.4 165 10 Q9SP16
 984 6 1.4 165 13 P87487
 985 6 1.4 165 16 Q9PIB5
 986 6 1.4 165 16 Q9A091
 987 6 1.4 166 1 Q93673
 988 6 1.4 166 3 Q60166
 989 6 1.4 166 5 Q9VK17
 990 6 1.4 166 16 Q08797
 991 6 1.4 167 4 Q96DC5
 992 6 1.4 167 8 Q9MNX1
 993 6 1.4 167 10 Q9SAS3
 994 6 1.4 167 16 Q67111
 995 6 1.4 167 16 Q825U9
 996 6 1.4 168 2 Q9EUC3
 997 6 1.4 168 2 Q85258
 998 6 1.4 168 2 Q85260
 999 6 1.4 168 2 Q85261
 1000 6 1.4 168 2 Q85262

Q9XG40 guillardia
 Q9YMT6 lymantria d
 Q9KI53 enterococcu
 Q12155 saccharomyc
 Q9NWT0 homo sapien
 Q8WZ09 homo sapien
 Q9BBL0 rhizobium l
 Q8R987 thermoanaer
 Q9KDR7 bacillus ha
 Q97Q11 streptococc
 Q9ZAR6 pseudomonas
 Q9BSC2 antillocapra
 Q9T9N9 terebratul
 Q9JVA7 neisseria m
 Q58287 pyrococcus
 Q02457 teladorsagi
 Q03961 oryza sativ
 Q9SP16 picea ruben
 P87487 oncorhynch
 Q9PIB5 campylobact
 Q9A091 streptococc
 Q93673 sulfolobus
 Q60166 schizosacch
 Q9VK17 drosophila
 Q08797 vibrio chol
 Q96DC5 homo sapien
 Q9MNX1 pagurus lon
 Q9SAS3 triticum ae
 Q67111 aquifex ae
 Q825U9 salmonella
 Q9EUC3 streptococc
 Q85258 streptococc
 Q85260 streptococc
 Q85261 streptococc
 Q85262 streptococc

ALIGNMENTS

RESULT 1
 Q25922 PRELIMINARY; PRT; 1720 AA.
 AC Q25922;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DE Precursor of the major merozoite surface antigens.
 DE Plasmodium falciparum (isolate NF54).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5843;
 RN [1]
 RP SEQUENCE OF 1069-1720 FROM N.A.
 RC STRAIN=NF54;
 RA Tolle R., Bujard H., Cooper J.A.;
 RL Exp. Parasitol. 0:0-0(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NF54;
 RA Tolle R.;
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NF54;
 RX MEDLINE=96123395; PubMed=8577332;
 RA Pan W., Tolle R., Bujard H.;
 RT "A direct and rapid sequencing strategy for the Plasmodium falciparum antigen gene gp190/MS1".
 RL Mol. Biochem. Parasitol. 73:241-244(1995).
 DR EMBL; Z35327; CAA84556.1; -
 DR InterPro; IPR000561; EGF-like.
 DR Pfam; PF00008; EGF; 1.
 KW Merozoite; Signal.

RESULT 2
 Q25923 PRELIMINARY; PRT; 652 AA.
 AC Q25923;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DE Merozoite surface antigen 1 (Fragment).
 DE Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE OF 1-298 FROM N.A.
 RC STRAIN=FCH5/NF7;
 RX MEDLINE=88142999; PubMed=2449612;
 RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G., Brown G.V., Anders R.F., Kemp D.J.;
 RT "Variation in the precursor to the major merozoite surface antigens of Plasmodium falciparum".
 RL Mol. Biochem. Parasitol. 27:291-302(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FCH5/NF7;
 RA Tolle R.;
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FCH5/NF7;
 RX MEDLINE=95354793; PubMed=7628566;
 RA Tolle R., Bujard H., Cooper J.A.;
 RT "Plasmodium falciparum: variations within the C-terminal region of merozoite surface antigen-1".
 RL Exp. Parasitol. 81:47-54(1995).
 DR EMBL; Z35328; CAA84557.1; -
 DR InterPro; IPR000561; EGF-like.
 DR Pfam; PF00008; EGF; 1.
 FT NON_TER 1 1

FT SIGNAL 1 19 POTENTIAL.
 SQ SEQUENCE 1720 AA; 195725 MW; 717B2FD1E637A8A3 CRC64;
 Query Match 86.88; Score 374; DB 5; Length 1720;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 58 AISVTMDNLSLGGFENEYDVLYLKLPLAGYRSLKQIEKNFTFNLNDILNSRLKRRKY 117
 DB 1327 AISVTMDNLSLGGFENEYDVLYLKLPLAGYRSLKQIEKNFTFNLNDILNSRLKRRKY 1386
 QY 118 FLDVLESDLAQFHRHISSEYIIEDSFKLNSQKNTLLSKYIKESVENDIKFAQEGIS 177
 DB 1387 FLDVLESDLAQFHRHISSEYIIEDSFKLNSQKNTLLSKYIKESVENDIKFAQEGIS 1446
 QY 178 YYEKVLAKYKDDLESIKVKIKEEKEKFPSPPTPPSPAKTDEQKESKELPFTLNITEL 237
 DB 1447 YYEKVLAKYKDDLESIKVKIKEEKEKFPSPPTPPSPAKTDEQKESKELPFTLNITEL 1506
 QY 238 YNVLNKKIDDYLYNLKAKINDCNVKEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIKK 297
 DB 1507 YNVLNKKIDDYLYNLKAKINDCNVKEKDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIKK 1566
 QY 298 LINDDTKMDLGLKLLTGLVQNPNTIISKLEGFQDMLNISQHCVKKQCPNSGCFR 357
 DB 1567 LINDDTKMDLGLKLLTGLVQNPNTIISKLEGFQDMLNISQHCVKKQCPNSGCFR 1626
 QY 358 HLDERECKLLNKKQEGDKCVENPNPTCNENNGCCDADATCTEEDSGSRKKITCECTK 417
 DB 1627 HLDERECKLLNKKQEGDKCVENPNPTCNENNGCCDADATCTEEDSGSRKKITCECTK 1686
 QY 418 PDSYPLFDGIFCSS 431
 DB 1687 PDSYPLFDGIFCSS 1700

```
SQ SEQUENCE 652 AA: 74292 MW: 2B6A87737B490A62 CRC64;
Query Match 79.8%; Score 344; DB 5; Length 652;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 344; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISVTMDNLSGFENEDYIYKPLAGVYRSKKQIEKNIPTFNLNDILNSRLKRRKY 117
DB 181 YNNLVNKIDYLYNLKAKINDCNVDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIKK 240

QY 259 AISVTMDNLSGFENEDYIYKPLAGVYRSKKQIEKNIPTFNLNDILNSRLKRRKY 318
DB 181 YNNLVNKIDYLYNLKAKINDCNVDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIKK 240

QY 118 FLDVLESMDLQPKHSSNEIYEDSKLNSQKNTLLSKYKIKESVENDIKFAEGIS 177
DB 319 FLDVLESMDLQPKHSSNEIYEDSKLNSQKNTLLSKYKIKESVENDIKFAEGIS 378

QY 178 YVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTTPSPAKTDQKKESKFLPLTNIETL 237
DB 379 YVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTTPSPAKTDQKKESKFLPLTNIETL 438

QY 238 YNNLVNKIDYLYNLKAKINDCNVDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIKK 297
DB 439 YNNLVNKIDYLYNLKAKINDCNVDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIKK 498

QY 298 LINDDTKKDMGLKLLSTGLVQNPNTIISKLEGGKFDMLNLSOHOCVKKQCPNSGCFR 357
DB 499 LINDDTKKDMGLKLLSTGLVQNPNTIISKLEGGKFDMLNLSOHOCVKKQCPNSGCFR 558

QY 358 HLDERECKLLNYKQEGDKVCNPNPTCNENNGGCDADATCTE 401
DB 559 HLDERECKLLNYKQEGDKVCNPNPTCNENNGGCDADATCTE 502

RESULT 3
ID Q25723 PRELIMINARY; PRT; 373 AA.
AC Q25723;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KENTA-1;
RX MEDLINE=98319411; PubMed=9657329;
RA Qari S.H., Shi Y.P., Goldman I.F., Nahlen B.L., Tibayrenc M., Lal A.A.;
RT "predicted and observed alleles of Plasmodium falciparum merozoite surface protein-1 (MSP-1), a potential malaria vaccine antigen.";
RL Mol. Biochem. Parasitol. 92:241-252(1998).
DR EMBL; AF040567; AAC39097.1; -
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1 373
FT NON_TER 373 373
SQ SEQUENCE 373 AA: 42873 MW: 886CF169A7AF5194 CRC64;
Query Match 53.6%; Score 231; DB 5; Length 373;
Best Local Similarity 100.0%; Pred. No. 3.1e-232;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISVTMDNLSGFENEDYIYKPLAGVYRSKKQIEKNIPTFNLNDILNSRLKRRKY 117
DB 1 AISTVTMDNLSGFENEDYIYKPLAGVYRSKKQIEKNIPTFNLNDILNSRLKRRKY 60

QY 118 FLDVLESMDLQPKHSSNEIYEDSKLNSQKNTLLSKYKIKESVENDIKFAEGIS 177
DB 61 FLDVLESMDLQPKHSSNEIYEDSKLNSQKNTLLSKYKIKESVENDIKFAEGIS 120

QY 178 YVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTTPSPAKTDQKKESKFLPLTNIETL 237
DB 121 YVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTTPSPAKTDQKKESKFLPLTNIETL 180

QY 238 YNNLVNKIDYLYNLKAKINDCNVDEAHVKITKLSDLKAIDDKIDLFKN 288
DB 181 YNNLVNKIDYLYNLKAKINDCNVDEAHVKITKLSDLKAIDDKIDLFKN 231

RESULT 5
ID Q43996 PRELIMINARY; PRT; 373 AA.
AC Q43996;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Merozoite surface protein-1 (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KENTA-2;
```

```
QY 238 YNNLVNKIDYLYNLKAKINDCNVDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIKK 297
DB 181 YNNLVNKIDYLYNLKAKINDCNVDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIKK 240

QY 298 LINDDTKKDMGLKLLSTGLVQNPNTIISKLEGGKFDMLNLSOHOCVKKQCP 350
DB 241 LINDDTKKDMGLKLLSTGLVQNPNTIISKLEGGKFDMLNLSOHOCVKKQCP 293

RESULT 4
ID Q43995 PRELIMINARY; PRT; 373 AA.
AC Q43995;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Merozoite surface protein-1 (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KENTA-1;
RX MEDLINE=98319411; PubMed=9657329;
RA Qari S.H., Shi Y.P., Goldman I.F., Nahlen B.L., Tibayrenc M., Lal A.A.;
RT "predicted and observed alleles of Plasmodium falciparum merozoite surface protein-1 (MSP-1), a potential malaria vaccine antigen.";
RL Mol. Biochem. Parasitol. 92:241-252(1998).
DR EMBL; AF040567; AAC39097.1; -
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1 373
FT NON_TER 373 373
SQ SEQUENCE 373 AA: 42873 MW: 886CF169A7AF5194 CRC64;
Query Match 53.6%; Score 231; DB 5; Length 373;
Best Local Similarity 100.0%; Pred. No. 3.1e-232;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISVTMDNLSGFENEDYIYKPLAGVYRSKKQIEKNIPTFNLNDILNSRLKRRKY 117
DB 1 AISTVTMDNLSGFENEDYIYKPLAGVYRSKKQIEKNIPTFNLNDILNSRLKRRKY 60

QY 118 FLDVLESMDLQPKHSSNEIYEDSKLNSQKNTLLSKYKIKESVENDIKFAEGIS 177
DB 61 FLDVLESMDLQPKHSSNEIYEDSKLNSQKNTLLSKYKIKESVENDIKFAEGIS 120

QY 178 YVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTTPSPAKTDQKKESKFLPLTNIETL 237
DB 121 YVEKVLAKYKDDLESIKKVIKEKEKFPSPPTTTPSPAKTDQKKESKFLPLTNIETL 180

QY 238 YNNLVNKIDYLYNLKAKINDCNVDEAHVKITKLSDLKAIDDKIDLFKN 288
DB 181 YNNLVNKIDYLYNLKAKINDCNVDEAHVKITKLSDLKAIDDKIDLFKN 231

RESULT 5
ID Q43996 PRELIMINARY; PRT; 373 AA.
AC Q43996;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Merozoite surface protein-1 (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KENTA-2;
```



```

RX MEDLINE=88011243; PubMed=3079521;
RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT "Allelic Dimorphism in a Surface Antigen Gene of the Malaria Parasite
RT Plasmodium falciparum.";
RL J. Mol. Biol. 195:273-287(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13343; BAA02604.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 570 AA; 64632 MW; 8674DEC89B2D562A CRC64;

Query Match 44.1%; Score 190; DB 5; Length 570;
Best Local Similarity 100.0%; Pred. No. 2.9e-189;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TPNLNLNDILNSRLKRRKYFLDVLESFLMFKHSSNEVYIEDSFKLLNSEQKNTLLKSY 158
Db 218 TPNLNLNDILNSRLKRRKYFLDVLESFLMFKHSSNEVYIEDSFKLLNSEQKNTLLKSY 277

QY 159 KYIKESVENDIRFAQEGISYYEKVLAKYKDDLESIKKVIKEEKKEKPPSPPTPPSPAKT 218
Db 278 KYIKESVENDIRFAQEGISYYEKVLAKYKDDLESIKKVIKEEKKEKPPSPPTPPSPAKT 337

QY 219 DEQKESKFLPFLTNTETLYNNLVNKIDYDYLINLAKINDCNVKEDEAHVKITKLSDLKA 278
Db 338 DEQKESKFLPFLTNTETLYNNLVNKIDYDYLINLAKINDCNVKEDEAHVKITKLSDLKA 397

QY 279 IDDKIDLFKN 288
Db 398 IDDKIDLFKN 407

RESULT 9
Q25968 ID Q25968 PRELIMINARY; PRT; 570 AA.
AC Q25968;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13345; BAA02606.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 570 AA; 64632 MW; 424BF553CCC2F2BE CRC64;

Query Match 44.1%; Score 190; DB 5; Length 570;
Best Local Similarity 100.0%; Pred. No. 2.9e-189;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TPNLNLNDILNSRLKRRKYFLDVLESFLMFKHSSNEVYIEDSFKLLNSEQKNTLLKSY 158

```

```

Db 218 TPNLNLNDILNSRLKRRKYFLDVLESFLMFKHSSNEVYIEDSFKLLNSEQKNTLLKSY 277

QY 159 KYIKESVENDIRFAQEGISYYEKVLAKYKDDLESIKKVIKEEKKEKPPSPPTPPSPAKT 218
Db 278 KYIKESVENDIRFAQEGISYYEKVLAKYKDDLESIKKVIKEEKKEKPPSPPTPPSPAKT 337

QY 219 DEQKESKFLPFLTNTETLYNNLVNKIDYDYLINLAKINDCNVKEDEAHVKITKLSDLKA 278
Db 338 DEQKESKFLPFLTNTETLYNNLVNKIDYDYLINLAKINDCNVKEDEAHVKITKLSDLKA 397

QY 279 IDDKIDLFKN 288
Db 398 IDDKIDLFKN 407

RESULT 10
Q25721 ID Q25721 PRELIMINARY; PRT; 373 AA.
AC Q25721;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povea M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
RT (MSP-1) in field-derived Plasmodium falciparum.";
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20726; AAA62217.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1
FT NON_TER 373
SQ SEQUENCE 373 AA; 42848 MW; EE0700233D7F4D4E CRC64;

Query Match 43.9%; Score 189; DB 5; Length 373;
Best Local Similarity 100.0%; Pred. No. 2.2e-186;
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISVTMDNLTSGFENEVDVYLYKPLAGVYRSKKQIEKNIFTNLTNLTNSRLKRRKY 117
Db 1 AISVTMDNLTSGFENEVDVYLYKPLAGVYRSKKQIEKNIFTNLTNLTNSRLKRRKY 60

QY 118 FLDVLESFLMFKHSSNEVYIEDSFKLLNSEQKNTLLKSYKIKESVENDIRFAQEGIS 177
Db 61 FLDVLESFLMFKHSSNEVYIEDSFKLLNSEQKNTLLKSYKIKESVENDIRFAQEGIS 120

QY 178 YYEKVLAKYKDDLESIKKVIKEEKKEKPPSPPTPPSPAKTDEQKESKFLPFLTNTETL 237
Db 121 YYEKVLAKYKDDLESIKKVIKEEKKEKPPSPPTPPSPAKTDEQKESKFLPFLTNTETL 180

QY 238 YNNLVNKID 246
Db 181 YNNLVNKID 189

RESULT 11
Q25727 ID Q25727 PRELIMINARY; PRT; 373 AA.
AC Q25727;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.

```

```
OC Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
(MSP-1) in field-derived Plasmodium falciparum."
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20732; AAA62223.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR NON_TER 1 373
FT NON_TER 373
SQ SEQUENCE 373 AA; 42814 MW; A79966CF38CC405C CRC64;

Query Match 38.3%; Score 165; DB 5; Length 373;
Best Local Similarity 100.0%; Pred. No. 2.4e-163;
Matches 165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TFNLNLNDILNSRLKRRKRYFLDVLESQKHISSNEYIIEFSKLLNSEQKNTLLKSY 158
DB 42 TFNLNLNDILNSRLKRRKRYFLDVLESQKHISSNEYIIEFSKLLNSEQKNTLLKSY 101

QY 159 KYIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKFPSSPTTPPSPAKT 218
DB 102 KYIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKFPSSPTTPPSPAKT 161

QY 219 DEQKESKFLPFLTNIETLYNNLVNKIDYLYNLKAKINDCNVEK 263
DB 162 DEQKESKFLPFLTNIETLYNNLVNKIDYLYNLKAKINDCNVEK 206

RESULT 12
Q25728 PRELIMINARY; PRT; 373 AA.
AC Q25728;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
(MSP-1) in field-derived Plasmodium falciparum."
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20733; AAA62224.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR NON_TER 1 373
FT NON_TER 373
SQ SEQUENCE 373 AA; 9CF4DDAF38CC4054 CRC64;

Query Match 38.3%; Score 165; DB 5; Length 373;
Best Local Similarity 100.0%; Pred. No. 2.4e-163;
Matches 165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 TFNLNLNDILNSRLKRRKRYFLDVLESQKHISSNEYIIEFSKLLNSEQKNTLLKSY 158
DB 42 TFNLNLNDILNSRLKRRKRYFLDVLESQKHISSNEYIIEFSKLLNSEQKNTLLKSY 101

QY 159 KYIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKFPSSPTTPPSPAKT 218
DB 102 KYIKESVENDIKFAQEGISYEKVLAKYKDDLESIKKVIKEEKFPSSPTTPPSPAKT 161

QY 219 DEQKESKFLPFLTNIETLYNNLVNKIDYLYNLKAKINDCNVEK 263
DB 162 DEQKESKFLPFLTNIETLYNNLVNKIDYLYNLKAKINDCNVEK 206
```

```
Db 162 DEQKESKFLPFLTNIETLYNNLVNKIDYLYNLKAKINDCNVEK 206

RESULT 13
Q43997 PRELIMINARY; PRT; 372 AA.
AC Q43997;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Merozoite surface protein-1 (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KENYA-3;
RX MEDLINE=98319411; PubMed=9657329;
RA Qari S.H., Shi Y.P., Goldman I.F., Nahlen B.L., Tibayrenc M.,
Lal A.A.;
RT "Predicted and observed alleles of Plasmodium falciparum merozoite
surface protein-1 (MSP-1), a potential malaria vaccine antigen."
RL Mol. Biochem. Parasitol. 92:241-252(1998).
DR EMBL; AF040369; AAC39099.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR NON_TER 1 372
FT NON_TER 372
SQ SEQUENCE 372 AA; 3ACD3AF6A6C047D53 CRC64;

Query Match 33.6%; Score 145; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.8e-142;
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISVTMDNLSGPFENEYDVLYLPLAGVYRSLLKQIEKNITFTNLNDILNSRLKRRKY 117
DB 1 AISVTMDNLSGPFENEYDVLYLPLAGVYRSLLKQIEKNITFTNLNDILNSRLKRRKY 60

QY 118 FLDVLESQKHISSNEYIIEFSKLLNSEQKNTLLKSYKESVENDIKFAQEGIS 177
DB 61 FLDVLESQKHISSNEYIIEFSKLLNSEQKNTLLKSYKESVENDIKFAQEGIS 120

QY 178 YEKVLAKYKDDLESIKKVIKEEKE 202
DB 121 YEKVLAKYKDDLESIKKVIKEEKE 145

RESULT 14
Q25725 PRELIMINARY; PRT; 372 AA.
AC Q25725;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
(MSP-1) in field-derived Plasmodium falciparum."
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20730; AAA62221.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR NON_TER 1 372
FT NON_TER 372
SQ SEQUENCE 372 AA; 42711 MW; E3DBB33F282CD5FC CRC64;
```

```
Query Match 33.6%; Score 145; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.8e-142;
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISVTMDNILSGFENEYDVIIKPLAGVYRSLKKQIEKNFTFNLNLDILNSRLKKRY 117
Db 1 AISVTMDNILSGFENEYDVIIKPLAGVYRSLKKQIEKNFTFNLNLDILNSRLKKRY 60

QY 118 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLSKYKIKESVENDIKFAOEGIS 177
Db 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLSKYKIKESVENDIKFAOEGIS 120

QY 178 YIEKVLAKYKDDLESIKKVIKEKE 202
Db 121 YIEKVLAKYKDDLESIKKVIKEKE 145

RESULT 15
Q25978 PRELIMINARY; PRT; 569 AA.
AC Q25978;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; DJ3351; BAA02612.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64536 MW; B8B9B63EECB1DA51 CRC64;

Query Match 33.6%; Score 145; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. No. 2.7e-142;
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISVTMDNILSGFENEYDVIIKPLAGVYRSLKKQIEKNFTFNLNLDILNSRLKKRY 117
Db 177 AISVTMDNILSGFENEYDVIIKPLAGVYRSLKKQIEKNFTFNLNLDILNSRLKKRY 236

QY 118 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLSKYKIKESVENDIKFAOEGIS 177
Db 237 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLSKYKIKESVENDIKFAOEGIS 296

QY 178 YIEKVLAKYKDDLESIKKVIKEKE 202
Db 297 YIEKVLAKYKDDLESIKKVIKEKE 321

RESULT 16
Q25983 PRELIMINARY; PRT; 569 AA.
AC Q25983;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; DJ3351; BAA02612.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64536 MW; B8B9B63EECB1DA51 CRC64;

Query Match 33.6%; Score 145; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. No. 2.7e-142;
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISVTMDNILSGFENEYDVIIKPLAGVYRSLKKQIEKNFTFNLNLDILNSRLKKRY 117
Db 177 AISVTMDNILSGFENEYDVIIKPLAGVYRSLKKQIEKNFTFNLNLDILNSRLKKRY 236

QY 118 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLSKYKIKESVENDIKFAOEGIS 177
Db 237 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLSKYKIKESVENDIKFAOEGIS 296

QY 178 YIEKVLAKYKDDLESIKKVIKEKE 202
Db 297 YIEKVLAKYKDDLESIKKVIKEKE 321

RESULT 16
Q25983 PRELIMINARY; PRT; 569 AA.
AC Q25983;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; DJ3351; BAA02612.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64536 MW; B8B9B63EECB1DA51 CRC64;
```

```
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
major merozoite surface proteins (MSPI) of Plasmodium falciparum from
field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; DJ3351; BAA02612.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64491 MW; 57A6B62FF72CE885 CRC64;

Query Match 33.6%; Score 145; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. No. 2.7e-142;
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISVTMDNILSGFENEYDVIIKPLAGVYRSLKKQIEKNFTFNLNLDILNSRLKKRY 117
Db 177 AISVTMDNILSGFENEYDVIIKPLAGVYRSLKKQIEKNFTFNLNLDILNSRLKKRY 236

QY 118 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLSKYKIKESVENDIKFAOEGIS 177
Db 237 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLSKYKIKESVENDIKFAOEGIS 296

QY 178 YIEKVLAKYKDDLESIKKVIKEKE 202
Db 297 YIEKVLAKYKDDLESIKKVIKEKE 321

RESULT 17
Q25924 PRELIMINARY; PRT; 651 AA.
AC Q25924;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface antigen 1 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RO-71;
RX MEDLINE=92275047; PubMed=1592091;
RA Olafsson P., Matile H., Certa U.;
RT "Plasmodium falciparum: the repetitive MSA-1 surface protein of the
RO-71 isolate is recognized by mouse antibody against the
nonrepetitive repeat block of RO-33."
RL Exp. Parasitol. 74:381-389(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=RO-71;
RX MEDLINE=95354793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
merozoite surface antigen-1."
RL Exp. Parasitol. 81:47-54(1995).
DR EMBL; Z35329; CAA84598.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1
SQ SEQUENCE 651 AA; 74134 MW; AA21376699255150 CRC64;

Query Match 33.6%; Score 145; DB 5; Length 651;
Best Local Similarity 100.0%; Pred. No. 3e-142;
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 AISVTMDNILSGFENEYDVIIKPLAGVYRSLKKQIEKNFTFNLNLDILNSRLKKRY 117
|||||
```



```
Db 259 AISVTMDNILSGFENEYDVIVLKLPLAGVYRSLRKLQIEKNIFTFNLNLDILNLSRUKKRY 318
Qy 118 FLDVLESLDMQFKHSSNEYIIDSFKLLNSEQKNTLKSXYIKESVENDIKFAQEGIS 177
Db 319 FLDVLESLDMQFKHSSNEYIIDSFKLLNSEQKNTLKSXYIKESVENDIKFAQEGIS 378
Qy 178 YEKVLAKYKDDLESIKKVIKEKE 202
Db 379 YEKVLAKYKDDLESIKKVIKEKE 403

RESULT 18
Q25970
ID Q25970 PRELIMINARY; PRT; 569 AA.
AC Q25970;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13347; BAA02608.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64502 MW; 4D761FF472753142 CRC64;

Query Match 32.7%; Score 141; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. NO. 4e-138;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 DFEAIKKLINDTTKKDMLGKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCYKQCP 350
Db 409 DFEAIKKLINDTTKKDMLGKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCYKQCP 468
Qy 351 ENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENGGCDADATCTEEDSGSSRKK 410
Db 469 ENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENGGCDADATCTEEDSGSSRKK 528
Qy 411 ITCECTKPDSPYPLFDGIFCSS 431
Db 529 ITCECTKPDSPYPLFDGIFCSS 549

RESULT 19
Q25980
ID Q25980 PRELIMINARY; PRT; 569 AA.
AC Q25980;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13354; BAA02615.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64530 MW; 4D734580516D2142 CRC64;

Query Match 32.7%; Score 141; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. NO. 4e-138;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13353; BAA02614.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64536 MW; 8008861DBECCED8DC CRC64;

Query Match 32.7%; Score 141; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. NO. 4e-138;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 DFEAIKKLINDTTKKDMLGKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCYKQCP 350
Db 409 DFEAIKKLINDTTKKDMLGKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCYKQCP 468
Qy 351 ENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENGGCDADATCTEEDSGSSRKK 410
Db 469 ENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENGGCDADATCTEEDSGSSRKK 528
Qy 411 ITCECTKPDSPYPLFDGIFCSS 431
Db 529 ITCECTKPDSPYPLFDGIFCSS 549

RESULT 20
Q25982
ID Q25982 PRELIMINARY; PRT; 569 AA.
AC Q25982;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13354; BAA02615.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64530 MW; 4D734580516D2142 CRC64;

Query Match 32.7%; Score 141; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. NO. 4e-138;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 DFEAIKKLINDTTKKDMLGKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCYKQCP 350
Db 409 DFEAIKKLINDTTKKDMLGKLLSTGLVQNFNTIISKLEIEGFQDMLNISQHCYKQCP 468
Qy 351 ENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENGGCDADATCTEEDSGSSRKK 410
Db 469 ENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENGGCDADATCTEEDSGSSRKK 528
Qy 411 ITCECTKPDSPYPLFDGIFCSS 431
Db 529 ITCECTKPDSPYPLFDGIFCSS 549

RESULT 21
Q25717
ID Q25717 PRELIMINARY; PRT; 372 AA.
```

```
AC Q25717;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
  (MSP-1) in field-derived Plasmodium falciparum."
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20655; AA62215.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1
FT NON_TER 372
SQ SEQUENCE 372 AA; 42687 MW; 3F2BF1152598FB10 CRC64;

Query Match 32.5%; Score 140; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 3e-137;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKLLNDTQKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 350
DB 233 DFEAIKLLNDTQKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 292

QY 351 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 410
DB 293 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 352

QY 411 ITCECTKPDSPYPLFDGIFCS 430
DB 353 ITCECTKPDSPYPLFDGIFCS 372

Query Match 32.5%; Score 140; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 3e-137;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKLLNDTQKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 350
DB 233 DFEAIKLLNDTQKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 292

QY 351 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 410
DB 293 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 352

QY 411 ITCECTKPDSPYPLFDGIFCS 430
DB 353 ITCECTKPDSPYPLFDGIFCS 372

Query Match 32.5%; Score 140; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 3e-137;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKLLNDTQKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 350
DB 233 DFEAIKLLNDTQKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 292

QY 351 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 410
DB 293 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 352

QY 411 ITCECTKPDSPYPLFDGIFCS 430
DB 353 ITCECTKPDSPYPLFDGIFCS 372

Query Match 32.5%; Score 140; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 3e-137;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKLLNDTQKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 350
DB 233 DFEAIKLLNDTQKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 292
```

```
QY 351 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 410
DB 293 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 352

QY 411 ITCECTKPDSPYPLFDGIFCS 430
DB 353 ITCECTKPDSPYPLFDGIFCS 372

RESULT 23
Q25720 PRELIMINARY; PRT; 372 AA.
AC Q25720;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
  (MSP-1) in field-derived Plasmodium falciparum."
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20655; AA62215.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
FT NON_TER 1
FT NON_TER 372
SQ SEQUENCE 372 AA; 42686 MW; 583A33972DB05FB4 CRC64;

Query Match 32.5%; Score 140; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 3e-137;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKLLNDTQKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 350
DB 233 DFEAIKLLNDTQKDMGLKLLSTGLVQNPNTIISKLEGKFDQMLNISQHCVKKQCP 292

QY 351 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 410
DB 293 ENSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKK 352

QY 411 ITCECTKPDSPYPLFDGIFCS 430
DB 353 ITCECTKPDSPYPLFDGIFCS 372

RESULT 24
Q25726 PRELIMINARY; PRT; 372 AA.
AC Q25726;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1, 42 kDa C-terminal region (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INFECTED HUMAN BLOOD SAMPLE;
RA Shi Y.-P., Alpers M.P., Povoa M.M., Nahlen B.L., Oloo A.G., Lal A.A.;
RT "Sequence of the C-terminal region of merozoite surface protein-1
  (MSP-1) in field-derived Plasmodium falciparum."
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20731; AA62222.1; -.

```

```
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
FT NON_TER 1
FT 372
FT NON_TER 372
SQ SEQUENCE 372 AA; 42734 MW; 19844F60D7011E52 CRC64;

Query Match 31.8%; Score 137; DB 5; Length 372;
Best Local Similarity 100.0%; Pred. No. 4.1e-134;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 ILSGFENEYDVIYKPLAGVYRSUKKQIEKNIETFNLNLDILNSRLKRRKYFDLVLESD 125
Db 9 ILSGFENEYDVIYKPLAGVYRSUKKQIEKNIETFNLNLDILNSRLKRRKYFDLVLESD 68

QY 126 LMQPKHSSNEYIIEDSFKLNSQKNTLKSRYIKESVENDIKFAQEGISYYEKVLAK 185
Db 69 LMQPKHSSNEYIIEDSFKLNSQKNTLKSRYIKESVENDIKFAQEGISYYEKVLAK 128

QY 186 YKDDLESIKKVIKEKE 202
Db 129 YKDDLESIKKVIKEKE 145

RESULT 25
Q9TVG8 PRELIMINARY; PRT; 218 AA.
AC Q9TVG8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MSAL protein (Fragment).
GN MSAL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FIJ4, AND FIS1;
RX MEDLINE=99263453; PubMed=10329360;
RA Lalitha P.V., Malhotra P., Chattopadhyay, Chauhan V.S.;
RT "Plasmodium falciparum: Variations in the C-terminal Cysteine-Rich
Region of the Merozoite Surface Protein-1 in Field Samples among
Indian Isolates.";
RL Exp. Parasitol. 92:12-18(1999).
DR EMBL; Y10600; CAA71609.1; -.
DR EMBL; Y10599; CAA71608.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
FT NON_TER 1
FT NON_TER 218
SQ SEQUENCE 218 AA; 24558 MW; EBD6A08FED6290F0 CRC64;

Query Match 27.1%; Score 117; DB 5; Length 218;
Best Local Similarity 100.0%; Pred. No. 1.9e-113;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCYKKQCP 350
Db 58 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCYKKQCP 117

QY 351 ENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGCCDADA 407
Db 118 ENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGCCDADA 174

RESULT 26
Q25969 PRELIMINARY; PRT; 569 AA.
AC Q25969;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).

Query Match 24.8%; Score 107; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.2e-102;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCYKKQCP 350
Db 409 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCYKKQCP 468

RESULT 28
Q25974 PRELIMINARY; PRT; 569 AA.
AC Q25974;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
major merozoite surface proteins (MSPI) of Plasmodium falciparum from
field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13348; BAA02609.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;

Query Match 24.8%; Score 107; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.2e-102;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCYKKQCP 350
Db 409 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCYKKQCP 468

RESULT 28
Q25974 PRELIMINARY; PRT; 569 AA.
AC Q25974;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
major merozoite surface proteins (MSPI) of Plasmodium falciparum from
field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13348; BAA02609.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;
```

```
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
major merozoite surface proteins (MSPI) of Plasmodium falciparum from
field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13348; BAA02607.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64485 MW; A26C45914AF01396 CRC64;

Query Match 24.8%; Score 107; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.2e-102;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCYKKQCP 350
Db 409 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCYKKQCP 468

QY 351 ENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGCCDADA 397
Db 469 ENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGCCDADA 515
```

RESULT 27

```
Q25974 PRELIMINARY; PRT; 569 AA.
ID Q25974
AC Q25974;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
major merozoite surface proteins (MSPI) of Plasmodium falciparum from
field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13348; BAA02609.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;
```

Query Match 24.8%; Score 107; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.2e-102;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 291 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCYKKQCP 350
Db 409 DFEAIKKLINDTTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCYKKQCP 468

QY 351 ENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGCCDADA 397
Db 469 ENSGCFRHLDERECKCLLNKQEGDKCVENPNPTCNENNGCCDADA 515
```

RESULT 28

```
Q25975
ID Q25975 PRELIMINARY; PRT; 569 AA.
AC Q25975;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
major merozoite surface proteins (MSPI) of Plasmodium falciparum from
field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13349; BAA02610.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64459 MW; ED4284B2867C9703 CRC64;

Query Match 24.8%; Score 107; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.2e-102;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEATKKLNDTCKDMLGKLLSTGLVQNFPTTIISKLEGRFQDMLNLSQHCVKKQCP 350
Db 409 DFEATKKLNDTCKDMLGKLLSTGLVQNFPTTIISKLEGRFQDMLNLSQHCVKKQCP 468

QY 351 ENSGCFRHLDERECKCLLNYKQEGDKVCVENPNTCNENNGGCDADA 397
Db 469 ENSGCFRHLDERECKCLLNYKQEGDKVCVENPNTCNENNGGCDADA 515

RESULT 29
Q25977
ID Q25977 PRELIMINARY; PRT; 569 AA.
AC Q25977;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
major merozoite surface proteins (MSPI) of Plasmodium falciparum from
field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13350; BAA02611.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64485 MW; A26C45914AF01396 CRC64;

Query Match 24.8%; Score 107; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.2e-102;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEATKKLNDTCKDMLGKLLSTGLVQNFPTTIISKLEGRFQDMLNLSQHCVKKQCP 350
Db 409 DFEATKKLNDTCKDMLGKLLSTGLVQNFPTTIISKLEGRFQDMLNLSQHCVKKQCP 468

QY 351 ENSGCFRHLDERECKCLLNYKQEGDKVCVENPNTCNENNGGCDADA 397
Db 469 ENSGCFRHLDERECKCLLNYKQEGDKVCVENPNTCNENNGGCDADA 515

RESULT 31
Q25979
ID Q25979 PRELIMINARY; PRT; 569 AA.
AC Q25979;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
major merozoite surface proteins (MSPI) of Plasmodium falciparum from
field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13352; BAA02613.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;

Query Match 24.8%; Score 107; DB 5; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.2e-102;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 DFEATKKLNDTCKDMLGKLLSTGLVQNFPTTIISKLEGRFQDMLNLSQHCVKKQCP 350
Db 409 DFEATKKLNDTCKDMLGKLLSTGLVQNFPTTIISKLEGRFQDMLNLSQHCVKKQCP 468

QY 351 ENSGCFRHLDERECKCLLNYKQEGDKVCVENPNTCNENNGGCDADA 397
Db 469 ENSGCFRHLDERECKCLLNYKQEGDKVCVENPNTCNENNGGCDADA 515

RESULT 31
Q25981
ID Q25981 PRELIMINARY; PRT; 1694 AA.
AC Q25981;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Major merozoite surface antigen.
GN GPI95.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCCL/HN;
RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.;
RT "Molecular cloning and sequence analysis of major merozoite surface
antigen (gpi95) gene of Plasmodium falciparum isolate FCC1/HN."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF218248; AAF27526.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW Merozoite.
SQ SEQUENCE 1694 AA; 192767 MW; B51634A49E0F6728 CRC64;

Query Match 24.8%; Score 107; DB 5; Length 1694;
Best Local Similarity 100.0%; Pred. No. 3.3e-102;
```

Query Match 24.8%; Score 107; DB 5; Length 1704;

Q25967	PRELIMINARY;	PRT;	569 AA.
ID	Q25967		
AC	Q25967;		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)		
DE	Major merozoite surface protein (Fragment).		
DE	MSPI.		
GN	Plasmodium falciparum.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5833;		
OX	[1]		
RN	SEQUENCE FROM N.A.		
RP	MEDLINE=93295445; Pubmed=8515786;		
RX	MEDLINE=93295445; Tanabe K., Kanbara H.;		
RA	Jongwutiwes S., Tanabe K., Kanbara H.;		
RT	"Sequence conservation in the C-terminal part of the precursor		
RT	major merozoite surface proteins (MSPI) of Plasmodium falciparum		
RT	field isolates.";		
RL	Mol. Biochem. Parasitol. 59:95-100(1993).		
RL	EMBL: D13344; BAA02605.1; -.		
DR	InterPro: IPR000561; EGF-like.		
DR			

RX	MEDLINE-21349158; PubMed-11456319;
RA	Da Silveira L.A., Ribeiro W.L., Kirchgatter K., Wunderlich G.,
RA	Matsuoka H., Tanabe K., Ferreira M.U.;
RT	"Sequence diversity and linkage disequilibrium within the merozoite
RT	surface protein-1 (Msp-1) locus of Plasmodium falciparum: a
RT	longitudinal study in Brazil.";
RJ	J. Eukaryot. Microbiol. 48:433-439(2001).
RL	EMBL: AF290875; AAK82766.1; -.
DR	InterPro: IPR000561; EGF-like.
DR	Pfam: PF00008; EGF; 1
FT	NON_TER 1
FT	SEQUENCE 95 AA; 10668 MW; C7786C3BEDA322ED CRC64;
SQ	
	Query Match 20.2%; Score 87; DB 5; Length 95;
	Best Local Similarity 100.0%; Pred. No. 1.9e-82;
	Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps
QY	336 MLNISQHCVKVKPCPNSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENGGCA 395
Dd	1 MLNISQHCVKVKPCPNSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENGGCA 60
QY	396 DATCTEEDSGSRKKITCECTKPDSP 422
Dd	61 DATCTEEDSGSRKKITCECTKPDSP 87
RESULT 38	
Q9BJU5	PRELIMINARY; PRT; 77 AA.
ID	Q9BJU5
AC	Q9BJU5;
DT	01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	Merozoite surface protein 1 (Fragment).
OS	Plasmodium falciparum.
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON	NCB1_TaxID=5833;
OX	[1]
RP	SEQUENCE FROM N.A.
RA	Eisen D.P., Cloonan N., Baddeley A., Eri R., Saul A.;
RT	"Antigenic drift and immune selection acting on merozoite surface
RT	protein 1-19 and merozoite surface protein 2 in independent field
RT	isolates of Plasmodium falciparum.";
RL	Submitted (DEC-2000) to the EMBL/GenBank/DDJB databases.
DR	EMBL: AF329518; AK19337.1; -.
FT	NON_TER 1
FT	SEQUENCE 77 AA; 8597 MW; 4EA799FF4B469D4 CRC64;
SQ	
	Query Match 17.9%; Score 77; DB 5; Length 77;
	Best Local Similarity 100.0%; Pred. No. 4.3e-72;
	Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps
QY	349 CPENSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENGGCADATCTEEDSGSSR 408
Dd	1 CPENSGCFRHLDERECKCLLNKQEGDKVCVENPNPTCNENGGCADATCTEEDSGSSR 60
QY	409 KKITCECTKPDSPFLD 425
Dd	61 KKITCECTKPDSPFLD 77
RESULT 39	
Q9BJU4	PRELIMINARY; PRT; 77 AA.
ID	Q9BJU4
AC	Q9BJU4;
DT	01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	Merozoite surface protein 1 (Fragment).
OS	Plasmodium falciparum.
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

```
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Eisen D.P., Cloonan N., Baddeley A., Eri R., Saul A.;
RT "Antigenic drift and immune selection acting on merozoite surface
RT protein 1-19 and merozoite surface protein 2 in independent field
RT isolates of Plasmodium falciparum.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF329519; AAK19338.1; -
FT NON_TER 1
FT NON_TER 77
SQ SEQUENCE 77 AA; 8587 MW; 4EA799FF4B4E69D4 CRC64;

Query Match 17.9%; Score 77; DB 5; Length 77;
Best Local Similarity 100.0%; Pred. No. 4.3e-72;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 CPENSGCGRHLDERECKCLLYKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSSR 408
Db 1 CPENSGCGRHLDERECKCLLYKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSSR 60

QY 409 KKITCECTKPDSPFLFD 425
Db 61 KKITCECTKPDSPFLFD 77

RESULT 40
Q9BJU3 Q9BJU3 PRELIMINARY; PRT; 77 AA.
AC Q9BJU3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Eisen D.P., Cloonan N., Baddeley A., Eri R., Saul A.;
RT "Antigenic drift and immune selection acting on merozoite surface
RT protein 1-19 and merozoite surface protein 2 in independent field
RT isolates of Plasmodium falciparum.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF329520; AAK19339.1; -
FT NON_TER 1
FT NON_TER 77
SQ SEQUENCE 77 AA; 8587 MW; 4EA799FF4B4E69D4 CRC64;

Query Match 17.9%; Score 77; DB 5; Length 77;
Best Local Similarity 100.0%; Pred. No. 4.3e-72;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 CPENSGCGRHLDERECKCLLYKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSSR 408
Db 1 CPENSGCGRHLDERECKCLLYKQEGDKCVENPNTCNENNGGCDADATCTEEDSGSSR 60

QY 409 KKITCECTKPDSPFLFD 425
Db 61 KKITCECTKPDSPFLFD 77
```

Search completed: May 12, 2003, 10:28:14
Job time : 68 secs